AFIT/DS/ENG/96-09

# Maximum Likelihood Estimation of Exponentials in Unknown Colored Noise for Target Identification in Synthetic Aperture Radar Images

DISSERTATION

Matthew Peter Pepin
Major, USAF

AFIT/DS/ENG/96-09

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the Department of Defense or the U. S. Government

AFIT/DS/ENG/96-09

# Maximum Likelihood Estimation of Exponentials in Unknown Colored Noise for Target Identification in Synthetic Aperture Radar Images

DISSERTATION

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Matthew Peter Pepin, B.S.E.E., M.E.E.E.

Major, USAF

September, 1996

AFIT/DS/ENG/96-09

# Maximum Likelihood Estimation of Exponentials in Unknown Colored Noise for Target Identification in Synthetic Aperture Radar Images

Matthew P. Pepin, B.S.E.E., M.E.E.E

Major, USAF

Approved:

Joseph J. Sacchini — 27 Sep 96
Dr. Joseph J. Sacchini, Committee Chairman

Byron M Welsh — 10-4-96
Dr. Byron M. Welsh, Committee Member

27 SEP 96
Dr. Gregory T. Warhola, Committee Member

Martin P. DeSimio — 4 OCT 96
Dr. Martin P. DeSimio, Committee Member

Fred Garber — 27 Sep 96
Dr. Fred Garber, Committee Member

Shankar Mall — 10-4-96
Dr. Shankar Mall, Dean's Representative

Robert A. Calico, Jr.
Dr. Robert A. Calico, Jr., Dean

*Acknowledgements*

In the course of my studies I have had the good fortune to be aided by many faculty and students in the pursuit of my degree. Whether in stimulating technical discussions or in acts of sheer good will, it has been my great pleasure to work with the highly competent people at the Air Force Institute of Technology.

In particular, I would like to thank my advisors, Professor Michael Clark and Professor Joseph Sacchini for directing my graduate research. I particularly want to express my appreciation to Doctor Clark for providing a firm framework for my understanding of signal processing through our many hours of technical debate. His patient mentoring of my novice work provided me with an appreciation for the profession of engineering research. I want to express my gratitude to Doctor Sacchini for his insights in the applications and practicabilities of the research; Ed Ingham for many hours of enlightening discourse in our mutual research interests; and Professors Welsh, Warhola and Desimio for their continuing critique of my work.

I thank my wife, Fay, and our two AFIT newborn daughters, Arabella and Ginaveve for enduring countless hours of my sleep deprived being, and my single mindedness for study. They supported my education unconditionally. Lastly, I thank my parents for enabling my appreciation of, and opportunities for this education.

Matthew Peter Pepin

*Table of Contents*

*List of Figures*

*Abstract*

The accurate and computationally efficient estimation of signals in noise has long been a field of intense study. The signal present in natural processes is many times well modeled as the sum of real or complex exponential functions. The noise for computational simplicity is often assumed to be white or uncorrelated. There exist, however, many cases where noise is, in fact, correlated. Accurate and efficient estimates of the signal in these cases require that the noise correlation be taken into account. This is case for the specific application of interest in this dissertation, Synthetic Aperture Radar (SAR), whose images of objects may be modeled as the sum of two-dimensional complex exponentials (the electromagnetic scattering centers on the target).

The maximum likelihood estimate of the signal is often considered the best possible estimate of the signal. While many white and colored noise maximum likelihood estimates have been developed, efficient solutions to the estimation of one- and two-dimensional exponentials in unknown colored noise do not exist.

This dissertation develops techniques for estimating exponential signals in unknown colored noise. The Maximum Likelihood (ML) estimators of the exponential parameters are developed. Techniques are developed for one and two-dimensional exponentials, for both the deterministic and stochastic ML model. The techniques are applied to Synthetic Aperture Radar (SAR) data whose point scatterers are modeled as damped exponentials. These estimated scatterer locations (exponentials frequencies) are potential features for model-based target recognition.

The estimators developed in this dissertation may be applied with any parametrically modeled noise having a zero mean and a consistent estimator of the noise covariance matrix. ML techniques are developed for a single instance of data in colored noise which is modeled in one dimension as 1) stationary noise, 2) autoregressive (AR) noise and 3) autoregressive moving-average (ARMA) noise and in two dimensions as 1) stationary noise, and 2) white noise driving an exponential filter. The classical ML approach is used to solve for parameters which can be decoupled from the estimation problem. The remaining nonlinear optimization to find the exponential frequencies is then solved by extending white noise ML techniques to colored noise. In the case of deterministic ML, the computationally efficient, one and two-dimensional Iterative Quadratic Maximum Likelihood (IQML)

methods are extended to colored noise. In the case of stochastic ML, the one and two-dimensional Method of Direction Estimation (MODE) techniques are extended to colored noise. Simulations show that the techniques perform close to the Cramér-Rao bound when the model matches the observed noise.

Application to SAR data first requires that damped exponentials have not been distorted by SAR processing. Then, 1-D colored noise techniques provide better estimates at low model orders (number of exponentials) than white noise techniques. The 2-D techniques based on the colored noise model also more accurately model SAR data than existing 2-D white noise techniques. With an appropriate focusing technique and matching technique for the exponentials in each dimension, scatterers are located with high resolution in SAR images and colored noise techniques improve these location estimates.

# Maximum Likelihood Estimation of Exponentials in Unknown Colored Noise for Target Identification in Synthetic Aperture Radar Images

## *I. Introduction*

### *1.1 Background*

This dissertation presents several new techniques for the estimation of sinusoidal or exponential signals in the presence of unknown colored noise. They include techniques for estimation in one and two dimensions. These new techniques can be used to estimate the locations of electromagnetic scattering centers on objects viewed with Synthetic Aperture Radar (SAR). They are quite general and can be applied to any exponential estimation problem involving colored noise. The use of these estimation techniques although centered on the particular characteristics of SAR (a single instance of data, modeled as the sum of damped exponentials), are not limited by these characteristics, but can be easily extended to multiple data instances and undamped exponentials.

The use of image data to detect and identify targets is pervasive in Air Force operations. One particular sensor of importance due its stand-off and resolution capabilities is Synthetic Aperture Radar (SAR). A SAR image is formed from radar target returns collected over a flight profile that simulates a very large aperture antenna. This SAR image data is used by aircrew or ground observers to provide reconnaissance, targeting, or navigation information. Attempts to convert the SAR data to a reduced set of derived features are ongoing efforts. With this reduced set of data more sophisticated pattern recognition algorithms can be used to efficiently detect and identify targets. The processor required to create SAR images with sufficient resolution is a major cost in SAR systems. The image resolution and size required for target recognition can easily exceed the cost associated with human interpretation of the data. One means of avoiding the cost associated with additional resolution is to directly use the radar scattering data or complex phase history without generating a SAR image. Pattern recognition algorithms can then be applied directly to features extracted from the phase history thereby avoiding the cost of SAR imagery.

The most significant of these extracted features provide a direct correlation to the locations of electromagnetic scattering centers on the target. These locations can be estimated with great accuracy by parametrically modeling the SAR data. To efficiently estimate these SAR model

parameters, the model must conform to or fit the SAR data. Methods of fitting damped exponentials to SAR data can more accurately estimate scatterer locations when unknown background clutter or colored noise is also considered. The colored noise techniques developed in this research improve the accuracy in estimating exponential scatterer locations and thus increase the suitability of using exponential scatterer locations as pattern recognition features.

## *1.2 Dissertation Outline*

This dissertation is developed along the following lines. Chapter 2 introduces the problem solved by this dissertation, explores the previous work in this area, and describes the approach in this dissertation to solve the problem. Chapter 3 presents notation and mathematical preliminaries used as a basis for subsequent developments. Chapter 4 introduces the specific problem of parametric estimation of scatterers in SAR data. The assumptions that lead to the damped exponential model are examined. Some of the potential pitfalls that are associated with collecting, processing, and imaging data containing damped exponentials are presented. In Chapter 5, the parametric estimation of one-dimensional damped exponentials in colored noise is examined. Because of its robust statistical properties, the Maximum Likelihood (ML) estimation technique is applied to solve the estimation problem. This chapter also develops several new estimation techniques that model the colored noise as a stationary noise sequence, an autoregressive (AR) noise sequence, or an autoregressive moving average noise sequence (ARMA). Both the deterministic ML model, which models the exponential amplitude, frequency and phase as unknown constants, and the stochastic ML model, which models the amplitudes as random, are examined. Chapter 6 examines the exponential estimation problem in colored noise for two dimensions. The 1-D deterministic techniques developed in the previous chapter are extended to 2-D for the deterministic ML model. Computationally less intensive techniques that estimate each dimension independently are developed for the stochastic ML model. Chapter 7 comes full circle and applies the colored noise techniques developed to 1-D and 2-D SAR data. The efficiency of the colored noise techniques in fitting the model to the data is compared with the white noise ML techniques and other techniques such as overmodeling (fitting more exponentials than expected to the data and selecting those that best fit the data). The suitability of scatterer locations as a pattern recognition feature is also examined. Chapter 8 presents conclusions and some potential areas of future research.

### *1.3 Contributions*

The contributions in this dissertation include the solution of the 1-D maximum likelihood problem of estimating deterministic exponentials in unknown colored noise. Computationally efficient estimation techniques for several noise models are developed. Similar techniques are developed for the stochastic maximum likelihood problem. A new spectral model for 2-D noise is developed leading to new 2-D estimation techniques for deterministic and stochastic maximum likelihood. Additionally, a mathematical model for representing damped exponentials on an irregularly sampled grid is developed and leads to a computationally efficient method for interpolating or focusing SAR images containing damped exponentials.

# II. Problem and Approach

## 2.1 Problem

The general problem solved in this dissertation is most easily understood in terms of the array processing problem. In the array processing problem, the aspects of exponential estimation in several dimensions can be observed. In addition, because of the general nature of the models used in this dissertation, the approaches derived have applications in all areas where spectral estimation, system identification, or harmonic retrieval play a role. The parameters of any process, that can be modeled as the sum of exponential signals in unknown Gaussian noise, may be estimated with the techniques derived in this dissertation.

In the array processing problem, a set of sensors are placed at fixed locations in a wavefield that consists of a small number of waves as shown in Figure 1. These sensors then measure the amplitude and phase of the wavefield at those locations. When the propagation frequency of the waves is known, an array may be designed such that measurement data can be used to estimate the directions of propagation of the waves. These directions of arrival (DOAs) are one of the parameters estimated by the techniques of this dissertation.

*2.1.1 Wavefield Model.* The wavefield model whose parameters are to be estimated is now constructed for the specific case of interest in this dissertation, the electromagnetic wavefield in free space. Applications to many other wavefields and medium exist, including for example, acoustic waves in the ocean (sonar) and acoustic waves in rock (seismography). The array processing problem for the electromagnetic wavefield is quickly derived from Maxwell's equations [14] [32]. The parametric model of the signal measured at a sensor, at time $t$, and location $\mathbf{r} = [\ r_x \quad r_y \quad r_z\ ]^T$ is

$$y(\mathbf{r},t) = s(\mathbf{r},t)e^{j(\omega t - \mathbf{r}^T\mathbf{k})} \tag{1}$$

where $s(\mathbf{r},t)$ is the amplitude of the wave, $\omega$ is the propagation frequency of the wave, and $\mathbf{k} = \frac{\omega}{c}[\ \cos\theta\sin\phi \quad \sin\theta\sin\phi \quad \cos\phi\ ]^T$ where $c$ is the speed of light. Now, with a set of appropriately placed sensors the wavefield can be measured, and the parameters $\{s,\theta,\phi,\omega\}$ estimated from the measured data. The position of the sensors samples the wavefield at various locations, $\mathbf{r}_m$ for $m = 0..M-1$, and the sensors are sampled with sampling period $T$, at discrete times, $t = nT$ for $n = 0..N-1$, to produce the measured data, $y(m,n)$. The simplest array geometry (sensor

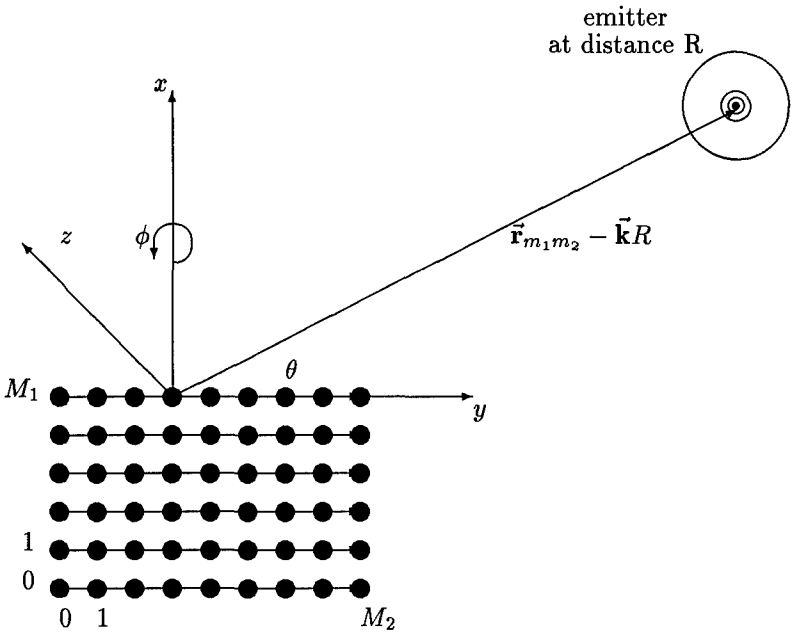


Figure 1. A Uniform Rectangular Array

positioning) is given by the uniform linear array (ULA). In the ULA, the sensors are placed in a line at a fixed interval. With the uniform linear array, only one direction of arrival angle, $\theta$, can be estimated. A simple extension of the ULA to two dimensions is the uniform rectangular array (URA), a rectangular grid of sensors $\mathbf{r}_{m_1m_2}$ for $m_1 = 0..M_1 - 1, m_2 = 0..M_2 - 1$, which allows both direction of arrival angles, $\theta$ and $\phi$, to be estimated. The array shown in Figure 1 is such an array.

*2.1.2 One-Dimensional and Two-Dimensional Data Characterizations.* Depending now on the statistical assumptions made about the parameters, the measured data may be characterized in one of several one- and two-dimensional estimation problems. Three of the most common characterization are detailed here. If sensor data measured a different times is considered uncorrelated, the data is temporally white and the sensor data for each time sample is considered a different statistical instance of the sensor data. In the 1-D ULA this leads to a one-dimensional estimation problem (parameters $\{s, \theta,$ known $\omega\}$) with the sensors $\mathbf{r}_m$ for $m = 0..M - 1$ where the data from each time sample is used to collect statistical information (mean, covariance) on the measured sensor data. With the same assumption, the 2-D uniform rectangular can be used to solve the two-dimensional estimation problem (parameters $\{s, \theta, \phi,$ known $\omega\}$) with sensors $\mathbf{r}_{m_1m_2}$ for $m_1 = 0..M_1 - 1, m_2 = 0..M_2 - 1$, and data instances $n = 0..N - 1$. This 2-D array processing problem may also be extended to synthetic aperture radar (SAR) as shown in Chapter 4. The last

data characterization considered in this dissertation is that involving a single sensor. This case reduces to a one-dimensional estimation problem (parameters $\{s,\omega\}$) with a single data instance, $m = 1$. This is a time-series analysis or one-dimensional spectral estimation problem.

As always in estimation problems, the goal of the estimation, accuracy, is limited by the available data. In many problems the numbers of sensors is extremely limited. A simple sonar array may contain less then ten sensors. The time available to estimate the parameters is also limited, both from signal duration and reaction time (i.e. as in a fire control system), and drives the computationally complexity of the estimation technique. Many fast estimation techniques have limited accuracy and make simplifying assumptions about the data for computationally tractable results. The problem then, succinctly stated, is to accurately estimate the parameters of particular array and data model chosen in a computationally efficient manner with as few assumptions on the data as possible. The extension of the array processing problem to the scatterer location problem in SAR data then allows the application of the array processing solutions to SAR data.

## *2.2 Previous Work*

This section details previous research that has been completed in the areas that this dissertation draws on as a basis for the new results herein. The areas covered include spectral estimation, maximum likelihood estimation, two-dimensional exponential estimation, array processing in colored noise, and scatterer identification in synthetic aperture radar images. As always, such a summary can not hope to be complete given the large body of work completed in each of the areas. Historical summaries of some of these areas are found in [35] [49] [66].

*2.2.1 Spectral Estimation.* The theory involved in the majority of the research in spectral estimation can be traced back to Fourier (1807). These nonparametric techniques are appropriately named Fourier analysis. It is, however, Bunson and Kirchhoff who are attributed with the idea that the spectrum of the light (signal) emitted by a substance can be used to characterize that substance and its physical properties [13]. The inherent utility of spectral estimation is the physical significance of the representation of a signal by its spectrum, whether the spectrum represents the molecular content of a substance, the key vibrations in an elastic body, the stability and response of a dynamic system, or the frequency and direction of propagating electromagnetic waves. Modern interest in characterizing the spectrum of random (statistical) signals begins with Wiener [83]

who developed the autocorrelation function for such signals and its Fourier transform relation with the power spectral density, the magnitude squared of the signal spectrum. Despite robust performance in most situations the Fourier techniques contain inherent limitations. The resolution of the Fourier spectrum is limited to a set of quantized frequencies and the variance of the estimated spectral energy at these frequencies is equal to the square of the estimate. Although a similar methodology for estimating exponential signals can be traced to Prony [55], the introduction of a parametric model by Yule [87] to characterize random processes heralded the introduction a high resolution, consistent estimators of the power spectrum of a random process. The basis for both these techniques is the homogeneous difference equation $\epsilon[n] = 0$,

$$a_p y[n] + a_{p-1} y[n-1] + \cdots + a_0 y[n-p] = \epsilon[n] \tag{2}$$

where the $a_i$ for $i = 0..p$ are the coefficients of the difference equation and $y[n]$ are samples of the random process. The solution to the difference equation is the sum of $p$ damped exponentials

$$y[n] = \sum_{i=1}^{p} s_i \lambda_i^n \tag{3}$$

where the $s_i$ are the exponential amplitudes and the $\lambda_i = e^{\alpha_i + j\beta_i}$ are the damped exponentials. With these parametric estimation techniques, the accuracy of the estimates is limited only by the noise present in the data. The key difference between the approaches of Prony and Yule is that Yule envisions the residuals remaining after applying the difference equation, $\epsilon[n]$, to the be a white noise driving force that is filtered by the difference equation to produce the colored noise process, $y[n], y[n-1], \cdots y[n-p]$. This is the first of the rational polynomial noise models, the autoregressive (AR) noise model. Prony, conversely, considers the exponentials to be fixed deterministic signals and the residuals an independent white noise process. Both methods arrive at the same spectral estimate thus they both may be used to estimate both exponential signals and colored noise. A key theorem that relates these two types of random processes is due to Wold [84] and states that any random process may be decomposed into two uncorrelated components, a completely deterministic process that may be exactly estimated from its past samples, and a random process that may be modeled as white noise driving a linear difference equation.

*2.2.1.1 Modern Spectral Estimation.* In modern spectral estimation the data is modeled as instances of a random vector, $\mathbf{y}_n = [\ y_{n,0} \quad y_{n,1} \quad \cdots \quad y_{n,M-1}]^T$ for $n = 0..N-1$, and the estimation of the parameters of the modeled data is accomplished by eigenanalysis of the covariance matrix (or some related matrix) of the data,

$$R_{\mathbf{yy}} = \mathrm{E}\{\mathbf{y}\mathbf{y}^*\}, \tag{4}$$

where $\mathrm{E}\{\cdot\}$ is the expected value operation and $(\cdot)^*$ indicates complex conjugate transpose. In the case where the data is modeled as a linear combination of signals and noise,

$$\mathbf{y} = G\mathbf{s} + \mathbf{w}, \tag{5}$$

where the columns of the matrix $G$ characterize the signals, $\mathbf{s}$ is a vector of the signal amplitudes, and $\mathbf{w}$ is zero-mean Gaussian noise. The model for the covariance matrix of the data is then

$$R_{\mathbf{yy}} = GPG^* + R_{\mathbf{ww}}, \tag{6}$$

where $P = \mathrm{E}\{\mathbf{s}\mathbf{s}^*\}$ and $R_{\mathbf{ww}}$ is the covariance matrix of the noise. The eigenanalysis methods, which begin with Pisarenko [54], were proven with Schmidt's multiple signal classification (MUSIC) method [61] which characterize the spectrum of the data in terms of signal and noise eigenspaces. For the case of white noise where $R_{\mathbf{ww}} = \sigma^2 I$ and $\sigma^2$ is the noise variance, this allows the noise portion of the data (the $M - p$ smallest eigenvalues) to be easily discarded. Other eigenanalysis methods included Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [57] which measures the signal with one set of sensors then measures the signal with a second identical but shifted (in location) set of sensors. An eigenanalysis of these two measurements produces an estimate of the signal. In the last type of eigenanalysis method, the principal components method [76], a form of Prony's method is used and the number of signals is overestimated. Thus, as noted earlier the estimate models both signal and noise. As in the previous methods, an eigenanalysis of the data allows the noise part of the measurement to be discarded. In the next section, the close relation between eigenanalysis method such as MUSIC and the maximum likelihood estimation is explored.

*2.2.2 Maximum Likelihood Estimation.* Maximum likelihood estimation is based on finding the probability density function (and model parameters) that best fits the observed data. In estimation problems, maximum likelihood (ML) estimation produces accurate estimators. Several notable large-sample properties of ML estimation have been proven. ML estimation produces consistent estimates (i.e. as the number of data samples grows the value of the estimator approached the true value). ML estimation has the best possible efficiency. If an unbiased estimator exists that attains the estimation performance bound, the Cramér-Rao bound (see Appendices C and D), it is the ML estimator. The ML estimator is asymptotically efficient. As the number of data samples grows the variance of the estimate approaches the Cramér-Rao bound (CRB). Additionally, Monte Carlo simulations have shown that ML estimators perform well on small size data samples as well.

*2.2.3 Deterministic and Stochastic Maximum Likelihood.* The method of maximum likelihood was originated by Fisher (1920) whose Fisher information matrix characterizes the CRB. Contributions to the method were made by Cramér (1946), Rao (1946), and Wald (1944). The application of maximum likelihood estimation to the array processing problem can be traced to [3] [4] [26] [80]. Most notably maximum likelihood estimation has been applied with two distinct assumptions. The first assumption is that the signals are nonrandom, in which case the signal is the mean of the data

$$\mathrm{E}\{\mathbf{y}\} = G\mathbf{s}, \tag{7}$$

and the variance of the data is due solely to additive noise

$$R_{\mathbf{yy}} = R_{\mathbf{ww}}. \tag{8}$$

This case is denoted as deterministic maximum likelihood. The second assumption is that the signal (amplitudes) is random. Then, the data has zero mean and covariance

$$R_{\mathbf{yy}} = GPG^* + R_{\mathbf{ww}} \tag{9}$$

where $P = \mathrm{E}\{\mathbf{ss}^*\}$. This case is called stochastic maximum likelihood. In a classic series of papers Stoica and Nehorai [68], [71], [72] demonstrate that for a fixed number of sensors, $m = 0..M-1$, as the number of data instances grows, $N \rightarrow \infty$, stochastic ML produces asymptotically better

estimates than deterministic ML. Also, in the papers the MUSIC spectral estimation is shown to be asymptotically equivalent to deterministic ML [68]. Alternatively, in the case where $N = 1$ and the number of sensors grows, $M \rightarrow \infty$, (or, equivalently, $M = 1, N \rightarrow \infty$), deterministic ML and stochastic ML asymptotically produce the same results [79]. All of the ML techniques cited thus far require some form of search of the signal space to find the columns of the matrix $G$ that best fit the data. A much more computationally attractive approach applicable to the ULA or time series problem was proposed independently by Kumaresan, Scharf and Shaw [33], and Bressler and Macovski [6]. In this approach, the deterministic ML problem is solved by iterating a linearized solution to the problem. The deterministic ML solution in this case is a quadratic optimization. With this technique, constraints are easily imposed to attain solutions that are a restricted set of the damped exponential model (i.e. exponentials on unit circle in complex plane, real valued exponentials, or sine waves), and increase estimation accuracy [34] [62]. This technique is called Iterative Quadratic Maximum Likelihood (IQML). The eigenspace equivalent of this technique, Method of Direction Estimation (MODE), was developed by Stoica and Sharman [69]. With an appropriate weighting factor MODE achieves the same accuracy as stochastic ML as the number of data instances, grows $N \rightarrow \infty$ [70].

*2.2.4 Two-Dimensional Exponential Estimation.* Several two-dimensional exponential estimation techniques have recently been developed that do not require a computationally intensive search of the 2-D frequency plane. These techniques either perform a one-dimensional estimation once in each dimension and pair the results (a 1-D by 1-D method) or they simultaneously estimate the 2-D frequencies (a full 2-D method). The first technique by Rao and Kung [2] uses an eigenanalysis of the measure data and a technique similar to ESPRIT to attain the spatial frequencies of the exponentials in each dimension. These frequencies are then paired according to which combinations best match the amplitudes of the measure data. The second technique, Matrix Enhanced Matrix Pencil (MEMP) from [25] also exploits the principle of ESPRIT to find the spatial frequencies. This method also requires pairing, however, it efficiently estimates the 2-D exponential frequencies by forming an estimate of the 2-D covariance matrix. A technique called 2D Prony [59] is an extension of the 1-D Prony technique. This method uses a technique of overmodeling (estimating more frequencies than the data contains) similar to the principal components method and has a robust method of selecting the two-dimensional exponential frequency combinations that produce the

highest energy signal. The remaining two techniques are derived from the two maximum likelihood approaches to the problem, stochastic ML and deterministic ML. The stochastic ML technique, 2D MODE, was developed by Li and Stoica [38]. This method uses the 1-D MODE technique and treats each columns or row as a separate data instance when estimating the frequencies of the columns or rows. A full 2-D technique was developed by Clark [11] which follows the approach of the 1-D IQML method. This method characterizes the linear space orthogonal to linear space of the signal and also parameterizes the problem in terms of a known noise covariance matrix. Since the general maximum likelihood problem formulation is easily parameterized with the covariance matrix of the noise, the ML techniques are also the most readily modified for the case of unknown colored noise.

*2.2.5 Array Processing in Colored Noise.* The white noise assumption in the array processing problem is in some cases invalid. Although consistent estimates are still attained by white noise techniques in the presence of colored noise, additional estimation accuracy can be attained by incorporating the noise coloration in the model [78]. When the noise covariance matrix is known the most common colored noise technique is called whitening. In this method the covariance matrix of the data is modified to make the noise component uncorrelated, then a white noise technique is applied to the data. In most cases the effect of this whitening on the signal is not taken into account and the signal estimate is still not optimal. For the ULA or URA, the known colored noise methods developed by Clark [9] produce efficient estimates for the one- and two-dimensional deterministic ML problem.

For the case of unknown colored noise, several approximate maximum likelihood methods have been developed. Such methods developed by Le Cadre and Wax preform a search over the potential signal and noise eigenvalues to find those that best represent the data covariance matrix [36], [81]. The methods of Böhme [5] and Friedlander [17] characterize the noise covariance matrix as a linear combination of known matrices, then search for the ML signal and noise estimate. Other unknown colored noise techniques include the instrumental variable approach [68] which estimates the spatial noise using temporally uncorrelated data samples, and an iterative Toeplitz covariance matrix method [77] which estimates the ML mean (signal) and Toeplitz covariance matrix (stationary noise). The previous techniques all involve collecting multiple instances of data to form the data covariance matrix.

The first unknown colored noise method for the single data instance was developed by Kay and Nagesha [29], [47], [48]. This method is an approximation of the ML estimate, however, the single data instance problem is of special interest in the SAR scatterer location problem. Also, the exact solution to unknown colored noise problem for stochastic ML has been developed by Ye and Degroat [86]. Although it also involves a multidimensional search for the ML estimates, the solution provides insight into the relationship between white and colored noise techniques and how computationally efficient white noise techniques might be extended to colored noise.

*2.2.6 Scatterer Location in Synthetic Aperture Radar Images.* The application of exponential estimation techniques to synthetic aperture radar is quite recent. The first application involves enhancements to the images themselves. When a small set of scatterers is identified in a SAR image, a simplified and more appealing SAR image can be produced. To this end, Gupta applies a form of Yule's method in each dimension [18], and Hua, et al. apply the MEMP method [24]. These estimation techniques are seen to estimate the locations of the dominant scatterers in the SAR images. A foundation for the work in scatterer location was developed by Sacchini who develops estimates of the scatterer locations in SAR data of simple objects (e.g. flat plates) [58]. Several recent results have be published by Li involving the application of Fourier and model-based methods to computer generated SAR data. In these results the Fourier based methods are the most robust, and the performance of model-based methods is significantly degraded due to the significant amount non-exponential signal present in the SAR data [39], [40]. This degradation is also noted in [53].

## *2.3 Approach*

As noted above, data from the intended application, SAR, contains energy not well modeled as exponential signal. This dissertation investigates modeling this energy as unknown colored noise. The SAR scatterer location problem and the array processing problem are related, thus unknown colored noise solutions to the array processing problem are explored. Of the methods presented that solve the array processing problem, the maximum likelihood methods provide the most accurate and consistent estimates of the array processing model. These methods allow for use of a parameterized noise covariance matrix and thus can eliminate the limiting white noise assumption and provide more accurate solutions. The approach then begins with these methods.

Next, of the array processing characterizations, the one-dimensional time series (or, equivalently, for the ULA, the single data instance case) has been researched the longest and is the most well understood. Many signal and noise models exist for time series analysis and maximum likelihood estimates for these models are well developed. Since the intended application, SAR, involves diverging electromagnetic waves, the signal model is the damped exponential model and the noise models are taken from the set of rational polynomial models. As noted earlier, the deterministic and stochastic ML assumptions do produce different results, thus, unknown colored noise solutions under both assumptions are explored. The results are compared in terms of the intended application, a single data instance of damped exponentials in colored noise.

The intended application is two-dimensional, thus, 2-D techniques are investigated. The two-dimensional deterministic and stochastic ML methods also allow for use of a parameterized noise covariance matrix. These two-dimensional techniques are also well related to their one-dimensional counterparts, thus, the methodology employed in the 1-D colored noise techniques is easily extended to two dimensions. Additionally, the different two-dimensional assumptions of 2-D IQML (2-D exponentials at distinct frequencies in each dimension) and 2-D MODE (exponentials at any intersection of a 1-D by 1-D frequency grid) allow development of techniques with radically different computational and scatterer location performance.

The array processing problem and the SAR scatterer location problem are related through a data interpolation or focusing step. The unknown colored noise array processing techniques may be applied to SAR data after the SAR data is interpolated to a uniform rectangular grid. Improvements in scatterer location and the fit of colored noise models to the data are investigated. The effects of the focusing method, colored noise model, the deterministic or stochastic assumption, the 1-D by 1-D or full 2-D exponential assumption, and the 1-D by 1-D matching technique used are also investigated.

# *III. Mathematical Preliminaries*

This chapter presents some mathematics that provide a foundation for the developments of this dissertation. Many references exist for the following material on matrix algebra and random vectors. The principal references used in developing these sections are [27] [41] [51] [66].

## *3.1 Matrix Algebra in Signal Processing*

In signal processing as in many areas of engineering, matrices and linear algebra provide an important tool for solving problems. This section introduces some of mathematical concepts that are exploited in solving the problems explored in this dissertation. Let $A$ and $B$ be complex matrices then the following notation defines specific operations on these matrices.

$A^T$ is the transpose of $A$.
$A^*$ is the complex conjugate transpose of $A$.
$\text{vec}(A)$ is the vector formed by stacking the columns of $A$.
$A \otimes B$ is the Kronecker product of $A$ and $B$.
$A \odot B$ is the Hadamard (Schur) product of $A$ and $B$.
$A^+$ is the pseudoinverse of $A$.
$P_A$ is the projection onto the column space of $A$.

*3.1.1 Singular Value Decomposition.* One facet of matrix algebra that has become of increasing importance in solving complicated problems is the Singular Value Decomposition (SVD). In the SVD any complex valued $n \times m$ matrix, $H$, may be decomposed as the product of three matrices

$$H = USV^*, \tag{10}$$

where $(\cdot)^*$ indicates complex conjugate transpose. Here, the $n \times n$ matrix $U$, and the $m \times m$ matrix $V$ are unitary matrices ($UU^* = I$ or $U^{-1} = U^*$) containing the left and right singular vectors, respectively. $S$ is an $n \times m$ diagonal matrix whose unique real-valued elements called the singular values are commonly ordered according to decreasing magnitude. The matrices $U$ and $V$ are not unique, however, they provide a valuable insight into the range $(\mathbf{x} : \mathbf{x} = H\theta, \mathbf{x} \in \mathsf{R}_{n\times 1}, \theta \in \mathsf{R}_{m\times 1})$ and null space $(\theta : H\theta = \mathbf{0})$ of the matrix $H$. If the rank of $H$ (the number of basis elements for the range of $H$) is $r$, then $S$ contains $r$ nonzero real singular values and the first $r$ columns of $U$ form a basis for the range of $H$, and the last $m - r$ columns of $V$ form a basis for the null space of $H$.

Efficient low rank ($k < r$) approximations of $H$ can also be formed by setting to zero the smallest $r - k$ singular values in $S$.

*3.1.2 Cholesky Decomposition.* A special case of the SVD occurs when $H$ is Hermitian ($H = H^*$). Then $n = m$ and the eigenspace decomposition of $H$ is

$$H = VDV^*, \tag{11}$$

where $V$ contains the eigenvectors of $H$, and $D$ is diagonal with real eigenvalues. When $D$ contains no negative elements, $H$ is positive semidefinite, and the singular value decomposition and eigenspace decomposition of $H$ are the same with $U = V$ and $S = D$. When the diagonal of $D$ also contains no zero values, $H$ is positive definite and the Cholesky decomposition of $H$ is

$$H = VD^{\frac{1}{2}}D^{\frac{1}{2}}V^* = R^*Q^*QR = R^*R, \tag{12}$$

where $D^{\frac{1}{2}}V^* = QR$ is the QR decomposition of $D^{\frac{1}{2}}V^*$, where $Q$ is unitary and the $n \times n$ matrix $R$ is upper triangular.

*3.1.3 Quadratic Solutions.* Often the problem to be solved involves an optimization and SVD is again a useful tool. The recurrent optimization problem in this dissertation is the minimization of a quadratic form

$$\min_{\mathbf{y}} \mathbf{y}^*A^*A\mathbf{y} = \min_{\mathbf{y}} \|A\mathbf{y}\|_2^2 \tag{13}$$

where $\|\cdot\|_2$ indicates the 2-norm, $A$ is $n \times m$, and $\mathbf{y} = [\ y_0 \quad -\mathbf{x}^T\ ]^T$. When the solution is constrained such that $\|\mathbf{y}\|_2^2 = 1$, the solution is minimizer of the Rayliegh quotient

$$\min_{\mathbf{y}} \frac{\mathbf{y}^*A^*A\mathbf{y}}{\mathbf{y}^*\mathbf{y}} \tag{14}$$

and the minimizing $\mathbf{y}$, $\mathbf{y}_{\min}$, is the eigenvector associated with the minimum eigenvalue of $A^*A$. When the constraint is $y_0 = 1$ and $A^*A$ is positive definite then

$$\mathbf{y}_{\min} = \frac{(A^*A)^{-1}\mathbf{c}}{\mathbf{c}^*(A^*A)^{-1}\mathbf{c}}, \tag{15}$$

where $\mathbf{c} = [\ 1 \quad \mathbf{0}^T\ ]^T$. When $y_0 = 1$, and $A^*A$ is rank deficient by one, then restating the minimization leads to the least squares problem

$$\min_{\mathbf{x}} \begin{bmatrix} 1 \\ -\mathbf{x} \end{bmatrix}^* \begin{bmatrix} \mathbf{b} \mid \tilde{A} \end{bmatrix}^* \begin{bmatrix} \mathbf{b} \mid \tilde{A} \end{bmatrix} \begin{bmatrix} 1 \\ -\mathbf{x} \end{bmatrix} = \min_{\mathbf{x}} \|\mathbf{b} - \tilde{A}\mathbf{x}\|_2^2 \tag{16}$$

where $\mathbf{b}$ contains the first column of $A$, and $\tilde{A}$ contains the last $N-1$ columns of $A$. The solutions to this problem are the Moore-Penrose pseudoinverses

$$\mathbf{x}_{\min} = \tilde{A}^+\mathbf{b} = \begin{cases} (\tilde{A}^*\tilde{A})^{-1}\tilde{A}^*\mathbf{b} & m < n \\ \tilde{A}^*(\tilde{A}\tilde{A}^*)^{-1}\mathbf{b} & n < m \end{cases} \tag{17}$$

where $(\cdot)^+$ indicates pseudoinverse. The general pseudoinverse of $A$ is constructed with the SVD and includes the case where $A$ has neither full column or row rank. The SVD pseudoinverse is

$$A^+ = VD^+U^* \tag{18}$$

where $D^+$ is constructed by inverting the non-zero elements of $D$.

*3.1.4 Projections.* The pseudoinverse then leads to another type of matrix that simplifies complex solution formula, the projection. A projection may be constructed for the matrix $A$ as

$$P_A = \begin{cases} AA^+ & m < n \\ A^+A & n < m \end{cases}. \tag{19}$$

The eigenvalues of the projection matrix are zero or one, and the eigenvectors associated with the eigenvalue one form a basis for the range of $A$. Thus for $m < n$

$$P_A A = A. \tag{20}$$

Consequently, the projection is idempotent,

$$P_A^2 = P_A. \tag{21}$$

Another potential way to simplify an optimization is to perform the optimization in the space orthogonal to the original problem formulation. For example,

$$\max_{\mathbf{y}} \mathbf{y} P_A \mathbf{y} = \min_{\mathbf{y}} \mathbf{y}(I - P_A)\mathbf{y}. \tag{22}$$

When a matrix representation for the orthogonal space, $G$, can be found such that $A^*G = 0$ and the matrix $[\ A \quad G\ ]$ is full rank, the projection onto the orthogonal space can be constructed,

$$P_G = G(G^*G)^{-1}G^*, \tag{23}$$

and the projections form a resolution of the identity matrix,

$$P_G + P_A = I. \tag{24}$$

Thus,

$$\min_{\mathbf{y}} \mathbf{y}(I - P_A)\mathbf{y} = \min_{\mathbf{y}} \mathbf{y} P_G \mathbf{y}. \tag{25}$$

*3.1.5 Signal Processing Matrices.* Two structured matrices play a special role in signal processing and define orthogonal subspaces. When evenly spaced samples of a periodic function are collected and placed in a vector $\mathbf{y}$, they may be modeled with an $n \times p$ Vandermonde matrix of the periodic components, $G$, and a vector of the component amplitudes $\mathbf{s}$,

$$\mathbf{y} = \begin{bmatrix} 1 & \cdots & 1 \\ \lambda_1 & \cdots & \lambda_p \\ & & \\ \lambda_1^{n-1} & \cdots & \lambda_p^{n-1} \end{bmatrix} \mathbf{s} = G\mathbf{s}, \tag{26}$$

The matrix that defines the space orthogonal to the matrix $G$ is an $n - p \times n$ Toeplitz matrix (the elements along each diagonal are identical). This matrix $A$ contains the coefficients of a polynomial

$a(\lambda) = a_p\lambda^p + \ldots a_1\lambda^1 + a_0$ whose roots are $\{\lambda_1, \ldots, \lambda_p\}$,

$$A^* = \begin{bmatrix} a_0 & a_1 & \cdots & a_p & \mathbf{0} \\ & \ddots & \ddots & \ddots & \\ \mathbf{0} & a_0 & a_1 & \cdots & a_p \end{bmatrix}. \tag{27}$$

Thus $A^*G = 0$.

*3.1.6 Kronecker and Hadamard Products.* Another useful matrix algebra tool is the Kronecker product which is defined as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1l}B \\ a_{21}B & a_{22}B & \cdots & a_{2l}B \\ \vdots & \vdots & & \vdots \\ a_{k1}B & a_{k2}B & \cdots & a_{kl}B \end{bmatrix}, \tag{28}$$

where $A$ is a $k \times l$ matrix, $B$ is a $m \times n$ matrix, and $A \otimes B$ is a $mk \times nl$ matrix that contains blocks of the matrix $B$ individually multiplied by each element of $A$. With the Kronecker product, operations can be performed simultaneously on the rows and columns of a matrix. For example, in the matrix product $AXB^T$, $A$ and $B$ are separable operations that act on the rows and columns of the matrix $X$, respectively. When the columns of the matrix $AXB^T$ are stacked in the *vec* (vectorize) operation [41], the result is conveniently given in terms of the Kronecker product as

$$\text{vec}(AXB^T) = (B \otimes A)\text{vec}(X). \tag{29}$$

The Kronecker product has several useful properties

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD) \tag{30}$$

$$(A + B) \otimes (C + D) = (A \otimes C) + (A \otimes D) + (B \otimes C) + (B \otimes D) \tag{31}$$

$$\text{tr}(A \otimes B) = \text{tr}(A)\text{tr}(B) \tag{32}$$

and

$$(C \otimes D)^{op} = C^{op} \otimes D^{op} \tag{33}$$

where *op* includes transpose, conjugate transpose, pseudoinverse and inverse when the individual inverses exist. Lastly, the Hadamard (Schur) product of two matrices of the same size is defined as

$$A \odot B = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1l}b_{1l} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2l}b_{2l} \\ \vdots & \vdots & & \vdots \\ a_{k1}b_{k1} & a_{k2}b_{k2} & \cdots & a_{kl}b_{kl} \end{bmatrix}. \tag{34}$$

### *3.2 Random Vectors*

If the elements of the vector $\mathbf{x}$ are samples of a random process, then the distribution of the vector can be described with a probability density function (PDF). In the case where the samples of the length $N$, complex vector $\mathbf{x}$ are complex circular symmetric Gaussian, the PDF is

$$f(\mathbf{x}) = \frac{1}{\pi^N |\sigma^2 R|} e^{-\frac{1}{\sigma^2}(\mathbf{x}-\mu)^* R^{-1}(\mathbf{x}-\mu)} \tag{35}$$

where for the expected value operation, $\mathrm{E}\{\cdot\}$,

$$\mathrm{E}\{\mathbf{x}\} = \mu, \tag{36}$$

and

$$\mathrm{E}\{(\mathbf{x}-\mu)^*(\mathbf{x}-\mu)\} = \sigma^2 R. \tag{37}$$

Additionally, for complex data the real and imaginary parts are assumed to be independent with

$$\mathrm{E}\{(\mathbf{x}-\mu)^T(\mathbf{x}-\mu)\} = 0 \tag{38}$$

to attain a valid PDF. The matrix $R$ is the covariance matrix of the vector $\mathbf{x}$ and has several interesting properties. The matrix is Hermitian symmetric and positive definite. The covariance matrix $R$ may be estimated in several different ways. The most common method is by forming the

outer product of several instances of the vector,

$$\hat{R} = XX^*, \tag{39}$$

where $X$ contains the instances of $\mathbf{x}$, $X = [\ \mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_K\ ]$. This method provides a consistent estimate of $R$ as $K \to \infty$, and when $K > N$, the covariance matrix is full rank with probability 1. The estimated $R$, however is not positive definite. In the case where $\mathbf{x}$ is stationary, a positive definite, full rank, covariance matrix can be formed as a Toeplitz matrix containing the estimated autocorrelation sequence of $\mathbf{x}$, $\hat{r}[i]$ for $i = -(N-1) \ldots N-1$,

$$R = \begin{bmatrix} \hat{r}[0] & \hat{r}[-1] & \cdots & \hat{r}[1-N] \\ \hat{r}[1] & \ddots & \ddots & \vdots \\ \vdots & \ddots & & \hat{r}[-1] \\ \hat{r}[N-1] & & \hat{r}[1] & \hat{r}[0] \end{bmatrix}. \tag{40}$$

This is equivalent to letting

$$X = \begin{bmatrix} x[0] & x[1] & \cdots & x[N-1] & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & x[0] & x[1] & \cdots & x[N-1] & 0 \\ 0 & \cdots & 0 & x[0] & x[1] & \cdots & x[N-1] \end{bmatrix} \tag{41}$$

be a $N \times (2N-1)$ matrix in Equation 39. Now, since $R$ is positive definite, its Cholesky decomposition, $R = L^*L$, provides an efficient method for creating realizations of noise whose covariance is $R$. The noise sequence $\mathbf{x}$ whose covariance is R is constructed from $L$ as

$$\mathbf{x} = L^*\delta$$

where $\delta$ is vector whose elements are a unit variance white noise sequence.

*3.2.1 Autocorrelation Sequence and Power Spectral Density.* The autocorrelation sequence,

$$\hat{r}_{xx}[k] = \frac{1}{N} \sum_{n=0}^{N-k} x[n]x[n+k] \qquad k = -N+1 \ldots 0 \ldots N-1, \tag{42}$$

is efficiently computed via the Fast Fourier Transform (FFT) as the Discrete Fourier Transform (DFT) of the data. When FFT is zero-padded to length $2N-1$, it interpolates all $2N-1$ auto-correlation coefficients required to form the matrix $R$. This Fourier transform of the data can also be viewed a Fourier transform of samples of the Power Spectral Density (PSD) of the data,

$$P_{xx}(e^{j\omega}) = \left| \sum_{n=0}^{N} x[n] e^{-jn\omega} \right|^2 . \tag{43}$$

In this case, the data samples, $x[n]$ $n = 0 \ldots N-1$, are coefficients of a filter satisfying a linear difference equation. In the most general form, this linear difference equation is a function of both the input sequence and the output sequence. The sequence then satisfies an autoregressive moving average, ARMA$(o, q)$, difference equation,

$$\begin{aligned} b_0 x[n] \;\; = \;\; & -b_1 x[n-1] \; \ldots - b_o x[n-o] \\ & + c_0 \epsilon[n] + c_1 \epsilon[n-1] \; \ldots + c_q \epsilon[n-q], \end{aligned} \tag{44}$$

where $\epsilon[n]$ is the input sequence. This allows the PSD to be characterized as the rational function of two polynomials. The ARMA PSD is

$$P_{xx}(e^{j\omega}) = \frac{\left| \sum_{n=0}^{q} c_n e^{-jn\omega} \right|^2}{\left| \sum_{n=0}^{o} b_n e^{-jn\omega} \right|^2}. \tag{45}$$

A special case of this PSD occurs when the single coefficient the numerator is $c_0$. This produces the autoregressive (AR) PSD.

*3.2.2 Maximum Likelihood.* Methods of efficiently estimating the parameters in these models are of interest. If an efficient estimator exists, the Maximum Likelihood (ML) technique is guaranteed to find this estimator. In this case, the ML estimator is unbiased and attains the Cramér-Rao estimation bound (CRB) (see Appendices C and D). The ML estimator has also been shown to be asymptotically efficient and asymptotically normally distributed (for a large number of data samples, high signal-to-noise ratio, or large number of data instances).

In the maximum likelihood technique, the PDF of the data is evaluated at the observed data. Then, the unknown parameters of this PDF function are optimized to find values of the parameters which maximize the function. These values are the ML estimates of the parameters.

## *IV. Modeling of Synthetic Aperture Radar Data with Exponentials*

This research develops exponential estimation techniques that can be applied to data sets such as Synthetic Aperture Radar (SAR) data. As a foundation, the relationship between the damped exponential model and the SAR process is established here. When SAR data is collected under certain constraints, the data can be modeled with a damped exponential model. A spatial ($\mathbf{k} = [\ k_x \quad k_y\ ]^T$) and temporal frequency ($\omega$) domain analysis of SAR data collection shows that SAR data may be used directly in exponential estimation algorithms with the addition of an interpolation step. In the interpolation step, the $\mathbf{k}-\omega$ domain data or phase history is transformed from the sampling grid on which it was attained (polar for chamber data, or polar-linear for airborne data) to a rectangular grid. This interpolation step may be skipped provided the $\mathbf{k}-\omega$ domain data are collected on a grid that is very nearly rectangular. The error induced by this approximation is examined later. Exponential estimation techniques model SAR data directly. Thus, any modification of the data will degrade the estimates. The use of filter windows, zero-padding, or inappropriate means of fitting the $\mathbf{k}-\omega$ domain data to a rectangular grid significantly degrades the performance of exponential estimation techniques. Like exponential estimation, the interpolation step also involves approximating from measured data. Thus, interpolation limits estimation accuracy. A good interpolation step or an estimation step that does not require interpolation are of interest for accurately locating scattering centers.

### *4.1 SAR Data Collection*

SAR data collection is modeled here in a two-dimensional perspective as shown in Figure 2. Here airborne SAR data is collected by spotlighting a ground target location. For a linear flight path the collection geometry involves a line (the flight path) and a non-colinear point (the target). The data collection can be thought of as occurring entirely in the plane represented by the point and the line (the slant plane). The target signature is collected as if the target is viewed orthogonal to this plane. The following simplifying assumptions are made: 1) the aircraft flight path is linear, 2) the radar transmits and receives the target echo at the same point in space, and 3) the Doppler in the signal is negligible. This model can also be expanded to include deviations from the linear path to account for actual aircraft flight path, or a circular path (chamber data).

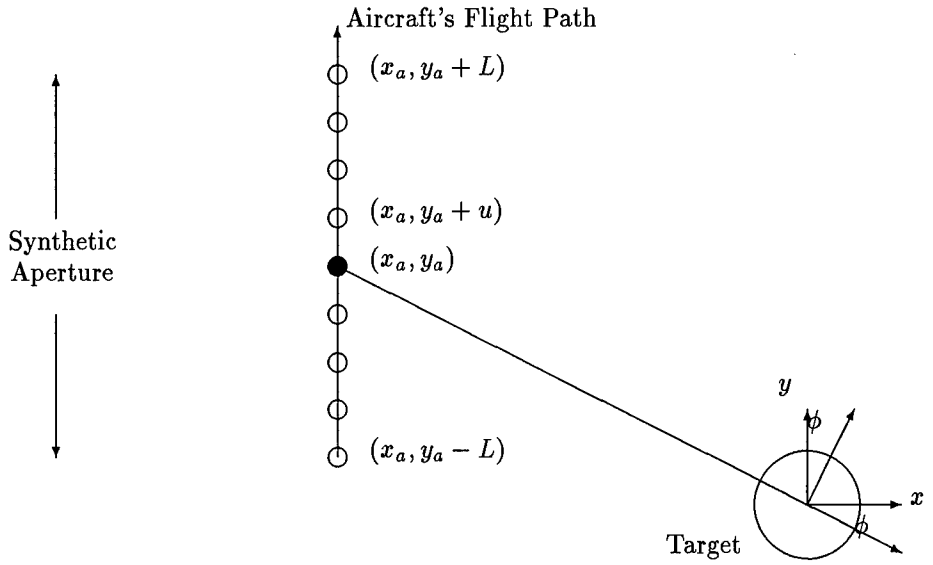


Figure 2. Spotlight SAR image formation geometry

With the above assumptions, and following the $\mathbf{k} - \omega$ domain analysis of [8] [63] [64] [65] the signal received by the radar at time $t$ and position $u$, $s(u,t)$, is the integral of the transmitted pulse, $p$, and the slant-plane target reflectivity function, $f(x,y)$, where $x$ and $y$ are coordinates in the slant plane. Then

$$s(u,t) = \int\int\int f(x,y,z)p(t - \frac{2}{c}\sqrt{(x-x_a)^2 + (y-u-y_a)^2})dxdydz, \tag{46}$$

where $c$ is the speed of light, $(x_a, y_a)$ is the center of the synthetic aperture, $z$ is the distance from the slant plane, and the limits of integration in this and all subsequent indefinite integrals are $\pm\infty$. If we take the Fourier transform with respect to $t$, defined as,

$$\mathcal{F}_t(s(u,t)) = \int_{-\infty}^{\infty} s(u,t)\exp(-j\omega t)dt, \tag{47}$$

and filter $s(u,t)$ with the matched filter for the pulse $p$, the signal becomes

$$P^*(\omega)S(u,\omega) = |P(\omega)|^2 \int\int\int f(x,y,z)\exp(-j2k\sqrt{(x-x_a)^2 + (y-u-y_a)^2})dxdydz, \tag{48}$$

where $k \equiv \frac{\omega}{c}$, and $P^*(\omega)$ is the response of the matched filter.

Next, assume that the target consists of point scatterers at distinct locations in space. Then

$$f(x,y,z) = \sum_n f_n \delta(x - x_n)\delta(y - y_n)\delta(z - z_n), \tag{49}$$

where the amplitude, $f_n$, includes target reflectivity, and a damping factor due to wave divergence which is inversely proportional to $\sqrt{(x_n - x_a)^2 + (y_n - u - y_a)^2}$. The image perspective orthogonal to the SAR collection plane includes scatterers at all depths, $z$, in the target, thus, the dependence on $z$ integrates out, giving a received signal

$$P^*(\omega)S(u,\omega) = |P(\omega)|^2 \int\int f(x,y)\exp(-j2k\sqrt{(x - x_a)^2 + (y - u - y_a)^2})dxdy, \tag{50}$$

where $f(x,y) = \sum_n f_n \delta(x - x_n)\delta(y - y_n)$.

The Fourier transform with respect to $u$ is given by a plane wave decomposition of the spherical wave, $\exp\left(-j2k\sqrt{(x - x_a)^2 + (y - u - y_a)^2}\right)$, which gives the Fourier transform relation [64]

$$\mathcal{F}_u\{\exp(-j2k\sqrt{x^2 + (y - u)^2})\} = \exp(-j\sqrt{4k^2 - k_u^2}x - jk_u y) \tag{51}$$

where the Fourier transform with respect to $u$, $\mathcal{F}_u$, is defined as,

$$\mathcal{F}_u(s(u)) = \int_{-\infty}^{\infty} s(u)\exp(-jk_u u)du. \tag{52}$$

Then, the $k_u - \omega$ spectrum of the received data is

$$\begin{aligned} P^*(\omega)S(k_u,\omega) &= |P(\omega)|^2 \int\int f(x,y)\exp(-j\sqrt{4k^2 - k_u^2}(x - x_a) - jk_u(y - y_a))dxdy \\ &= |P(\omega)|^2 \exp(j\sqrt{4k^2 - k_u^2}x_a + jk_u y_a) \int\int f(x,y)\exp(-j\sqrt{4k^2 - k_u^2}x - jk_u y)dxdy \\ &= |P(\omega)|^2 \exp(j\sqrt{4k^2 - k_u^2}x_a + jk_u y_a)F(\sqrt{4k^2 - k_u^2}, k_u), \end{aligned} \tag{53}$$

where $F$ is the Fourier transform of $f$. If we normalize the received signal by $1/|P(\omega)|^2$, it is simply the 2-D Fourier transform of the target reflectivity function $f(x,y)$ with a phase shift term to account for the target angle with respect to the aperture (squint). If samples of $F(\cdot,\cdot)$ were available at uniformly spaced intervals of $\sqrt{4k^2 - k_u^2}$ and $k_u$, $f(x,y)$ could be recovered error-free by this procedure [64]. The data collection geometry, however, does not permit this and

uniformly spaced samples must be interpolated from the available data. The most common form of interpolation is to assume that a point scatterer exists at each sample in the image, $f(x,y)$. This is called focused SAR.

Focused SAR for chamber data [42] involves phase shift terms to align the samples in lines perpendicular to the $k_x$ and $k_y$ axes.. The focused SAR formulation is

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi}\int_0^{\pi}\int_{-\infty}^{\infty} F(k,\theta)\exp(-j2kx\cos\theta - j2ky\sin\theta)\,k\,dkd\theta \\ &= \frac{K}{2\pi}\int_0^{2\pi}\int_{-\infty}^{\infty} F(2k,\theta)\exp(j2kx(1-\cos\theta)+jy(\theta-2k\sin\theta))\exp(-jx2k-jy\theta)2\,dkd\theta. \end{aligned} \tag{54}$$

Fourier transforming both sides of Equation 54 shows that the collected data $F(k,\theta)$ can be simply modulated by $\exp(j2kx(1-\cos\theta)+jy(\theta-2k\sin\theta))$ to attain the rectangularly sampled $F(k_x,k_y)$. In this method, the contribution of every measurement is added coherently with respect to each pixel, assuming that a point scatterer exists at every pixel. This focusing, however, degrades damped exponential signals in the data by convolving them with a point spread function, and limits the ability to attain high resolution estimates of the exponentials. More appropriate methods of interpolation are discussed in the following sections.

### *4.2 Fitting the Damped Exponential Model*

The point scatterer assumption in the SAR process leads directly to the exponential model with a damping factor, $f_n$, to account for wave divergence. With the point scatterer assumption, the Fourier transform of $f(x,y)$ is the 2-D damped exponential model

$$F(k_x,k_y) = \sum_n f_n \exp(-jk_x x_n)\exp(-jk_y y_n). \tag{55}$$

Notice that this it equivalent to the array processing model

$$F(\mathbf{k}) = \sum_n f_n \exp(-j\omega t - j\mathbf{k}^T\mathbf{x}_n) \quad t=0, \tag{56}$$

where $\mathbf{k}=[\ k_x \quad k_y\ ]^T$ and $\mathbf{x}_n=[\ x_n \quad y_n\ ]^T$ at time $t=0$. Thus, the array processing model for which many parameter estimation techniques have been developed can be exploited. The difficulty in applying this model lies in the sampling of the target reflectivity function.

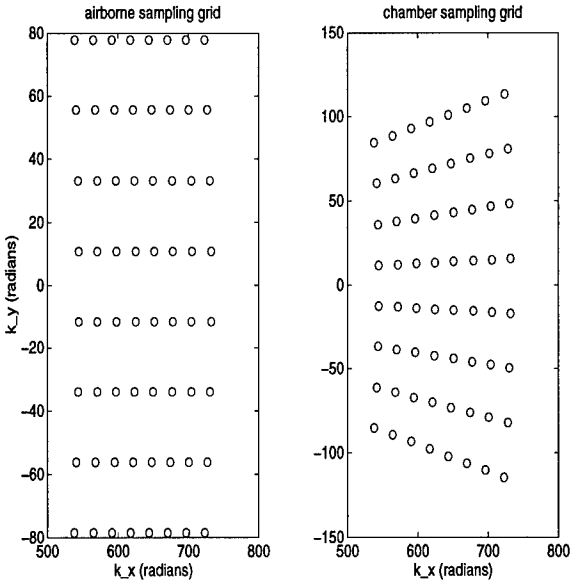


Figure 3. Target Spatial Frequency Samples

The calculation of the Fourier transform of the target reflectivity function (for the airborne case) is error free. However, when samples are taken uniformly in $u$, the samples are not uniform in $k_x$, and when the path is circular, the samples are on a polar grid in the $(k_x, k_y)$ plane. The sampling grid for the airborne and chamber cases are plotted in Figure 3. Interpolation must be performed between the data sampling grid and the rectangular grid needed to estimate damped exponential parameters. Several focusing methods that create SAR images do this. However, these assume that a point scatterer exists at each pixel and thus impose a point spread function on the data. This windowing degrades the high resolution estimates of parametric estimation methods [30]. Inappropriate means of interpolating a SAR image such as zero-padding X-PATCH data also destroys the performance of exponential estimation techniques [53]. Gradient methods also exist for focusing SAR data at existing scatters, but these techniques are computationally burdensome [15] [20]. This leaves three solutions: 1) collect data on an almost rectangular grid, 2) interpolate the data, 3) estimate exponentials from nonuniform data.

*4.2.1 Rectangular Assumption.* The rectangular assumption has been used in SAR techniques that incorporate exponential estimation [24] [58]. The error incurred by using the polar or polar-rectangular grid data directly is determined by examining the deviation of these grids from the rectangular grid (see Figure 4). For chamber data on the polar grid, the data are collected

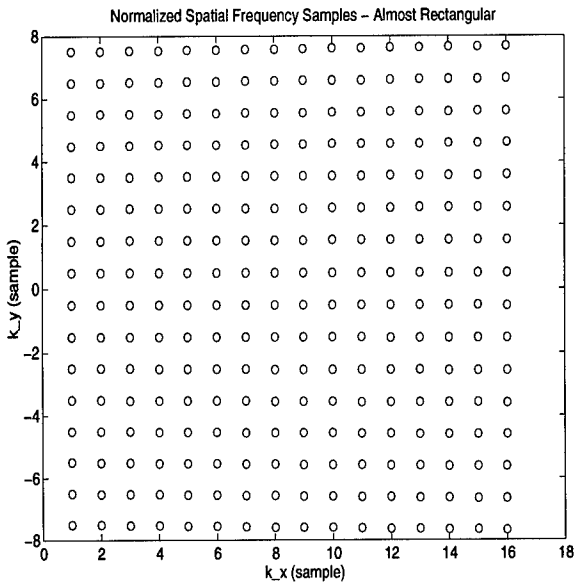


Figure 4. Normalized Target Spatial Frequency Samples for Chamber SAR. The small angular and frequency increments cause the grid to be almost rectangular.

about an angle, $\phi$, from the $x$-axis, in a beam of $\theta \in [-\theta_b, \theta_b]$, and at frequencies $k \in [k_{\min}, k_{\max}]$. The first requirement is to achieve better than Fourier resolution, thus, each sample, $F(k_{m_1}, \theta_{m_2})$, must lie in a separate cell in the rectangular grid. In Figure 4 if the first sample in a row is lower than the last sample of an adjacent row, then resolution of a scatterer to less than a pixel (Fourier bin) is not possible. This occurs first in Figure 4 in top row of samples when

$$k_{\min} < \frac{N-2}{N-1} k_{\max}. \tag{57}$$

The error in point scatterer location is given by the frequency estimation accuracy. This accuracy is bounded by the maximum $k_x$ or $k_y$ deviation of samples within a column or row of the rectangular grid. This error consists of two orthogonal components, $e_{k_x}$ and $e_{k_y}$. Since the polar data are circularly symmetric, $\phi$ is arbitrary and is chosen to be $0°$. The error in the $k_x$ direction is greatest at the outer ring and is

$$e_{k_x} = k_{\max} \cos 0° - k_{\max} \cos \theta_b = k_{\max}(1 - \cos \theta_b). \tag{58}$$

The error in the $y$ direction is the deviation of the radial lines from parallel beam center and is

$$e_{k_y} = k_{\max} \sin\theta_b - k_{\min} \sin\theta_b = (k_{\max} - k_{\min}) \sin\theta_b. \tag{59}$$

The maximum error is then the maximum of these errors. Airborne SAR is regularly sampled in the $k_y$ direction and thus contains only the $e_{k_x}$ error term.

*4.2.2 Interpolation.* Since the $\mathbf{k} - \omega$ domain data resembles the sum of complex exponentials the sampled data is quite smooth. Intermediate samples may be estimated from the polar sampled data to form a rectangular grid. The simplest form of interpolation, phase-shifting the data to a rectangular grid was described earlier. More complex methods of interpolation including 1-D and 2-D Lagrange interpolation polynomials [37] and an inverse distance method [60] are also examined.

*4.2.3 Scattering Center Focus.* This section introduces a new compact model for non-uniformly sampled 2-D exponentials. With this model a new computationally efficient method of focusing SAR phase history data by simply applying phase shifts to the data samples is developed.

The phase history data $F$, the Fourier transform of the SAR image, can be written as 2-D damped exponential data from $p$ point scatterers and noise $N$

$$F[m_1, m_2] = \sum_{i=1}^{p} s_i \lambda_i^{m_1} \gamma_i^{m_2} + N[m_1, m_2] \qquad \begin{array}{l} 0 \le m_1 \le M_1 - 1 \\ 0 \le m_2 \le M_2 - 1, \end{array} \tag{60}$$

where the $\lambda_i$ and $\gamma_i$ are the complex spatial frequencies, and the $s_i$ are the complex amplitudes of the scatterers and $N[m_1, m_2]$ is additive noise. Arranging $F$ in an $M_1 \times M_2$ matrix gives

$$F = GSH^T + W, \tag{61}$$

where

$$G = [\ \mathbf{g}_1 \quad \mathbf{g}_2 \quad \cdots \quad \mathbf{g}_p\ ] \quad \mathbf{g}_i = \Psi_{M_1}(\lambda_i),$$

$$S = \mathrm{diag}([\ s_1 \quad s_2 \quad \cdots \quad s_p\ ]),$$

$$H = [\ \mathbf{h}_1 \quad \mathbf{h}_2 \quad \cdots \quad \mathbf{h}_p\ ] \quad \mathbf{h}_i = \Psi_{M_2}(\gamma_i),$$

and

$$\Psi_M(z) = [\ 1 \quad z \quad z^2 \quad \cdots \quad z^M\ ]^T.$$

This model assumes that the data are sampled on a rectangular grid. A reformulation of Equation 61 will expose how the rectangular grid is involved in the model and how it may be replaced with a different grid to reflect the data collection. Consider a 2-D damped exponential with a single exponential in each dimension. The matrix form may then be expressed as

$$F = s_i \lambda_i^{\mathbf{X}} \odot \gamma_i^{\mathbf{Y}} \tag{62}$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 2 & 2 & \cdots & 2 \\ \vdots & \vdots & & \vdots \\ M_1 & M_1 & \cdots & M_1 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} 1 & 2 & \cdots & M_2 \\ 1 & 2 & \cdots & M_2 \\ \vdots & \vdots & & \vdots \\ 1 & 2 & \cdots & M_2 \end{bmatrix}, \tag{63}$$

and the scalar $\lambda_i$ taken to a matrix power results in a matrix of the same size with elements $\lambda_i^{x_{m_1 m_2}}$ where $x_{m_1 m_2}$ is the $m_1 m_2^{th}$ element of $\mathbf{X}$ and $\odot$ represents the element-wise Hadamard or Schur product of two matrices. $\mathbf{X}$ and $\mathbf{Y}$ represent the coordinates of the sampling grid on which the data is taken. To express a damped exponential which is sampled with a non-rectangular grid, simply replace the matrices $\mathbf{X}$ and $\mathbf{Y}$ with the rectangular coordinates of that grid. For the multiple exponential case,

$$F = \sum_{i=1}^{p} s_i \lambda_i^{\mathbf{X}} \odot \gamma_i^{\mathbf{Y}}. \tag{64}$$

Each element of $F$ is

$$\begin{aligned} F_{m_1 m_2} &= \sum_{i=1}^{p} s_i \lambda_i^{x_{m_1 m_2}} \gamma_i^{y_{m_1 m_2}} \\ &= \sum_{i=1}^{p} \left(s_i \lambda_i^{m_1} \gamma_i^{m_2}\right) \lambda_i^{x_{m_1 m_2} - m_1} \gamma_i^{y_{m_1 m_2} - m_2} \end{aligned}$$

The damped exponential model is distorted at each pixel by the weighting factors $\left\{\lambda_i^{x_{m_1 m_2} - m_1} \gamma_i^{y_{m_1 m_2} - m_2}\right\}_{i=1\ldots p}$. For conversion to a normalized polar grid these weights are $\lambda_i^{k_{m_1} \cos\theta_{m_2} - k_{m_1}} \gamma_i^{k_{m_1} \sin\theta_{m_2} - \theta_{m_2}}$ where $k_{m_1} = m_1 \triangle k$ and $\theta_{m_2} = m_2 \triangle \theta$. Assume that the $\lambda_i$ and $\gamma_i$ are undamped exponentials. Then, $\lambda_i^{x_{m_1 m_2} - m_1} \gamma_i^{y_{m_1 m_2} - m_2}$ simply applies a phase shift to each

term in $F_{m_1 m_2}$ and the average phase shift for equal amplitude scatterers is

$$\angle_{m_1 m_2} = \frac{1}{p} \sum_{i=1}^{p} \angle \lambda_i^{k_{m_1} \cos \theta_{m_2} - k_{m_1}} + \angle \gamma_i^{k_{m_1} \sin \theta_{m_2} - \theta_{m_2}}. \tag{65}$$

Thus, the focusing term in Equation 54 is clearly inappropriate since it assumes that a scattering center exists at every pixel in the image. This focusing distorts scattering center locations.

Equation 65 is easily implemented. Estimate the $\lambda_i$ and $\gamma_i$ by picking the brightest pixels in the image and calculating their spatial frequency. Any other 2-D frequency estimation technique can be used. However, since only the scattering center pixel locations are required, the DFT of $F(k_{m_1}, \theta_{m_2})$ provides the best result. This method also provided a fast means of improving SAR image quality when 2-D interpolation methods are too intensive. Figures 5, 6, and 7 show images resulting from the various focusing methods used on chamber data of a C-29 aircraft and the unfocused data (rectangular assumption).

*4.2.4 1-D and 2-D Interpolation.* For the 1-D Lagrange interpolation polynomial, the $k_y$ dimension for chamber data was interpolated. An $N^{th}$ order interpolation polynomial was formed for each frequency in the polar grid. This polynomial was then evaluated at the rectangular grid points giving the interpolated data. The most computationally intensive techniques examined were 2-D. The inverse distance method used the 2-D gradient to interpolate, and a Lagrange polynomial interpolated the stacked columns of the 2-D data. As a basic measure of the relative merit of each technique, an ideal image of point sources was sampled on an 'almost rectangular' polar grid, then interpolated by each technique to a rectangular grid. Figure 8 shows the energy loss of each interpolation method, $\bar{e} = \frac{\|F - \hat{F}\|_F}{\|F\|_F}$ where $\|\cdot\|_F$ indicates the Frobenius norm. As expected the 2-D interpolation methods show less degradation.

### *4.3 Results*

Accuracy is important in determining relative scatterer locations. When exponential techniques model un-interpolated data, minimizing the energy between the model and an image does not accurately locate the scatterers in an image. Thus, ideal data is used here to determine the performance of interpolation and estimation using SAR data. Ideal SAR data of two point scatterers was created to test the accuracy of several exponential estimation techniques with different interpo-

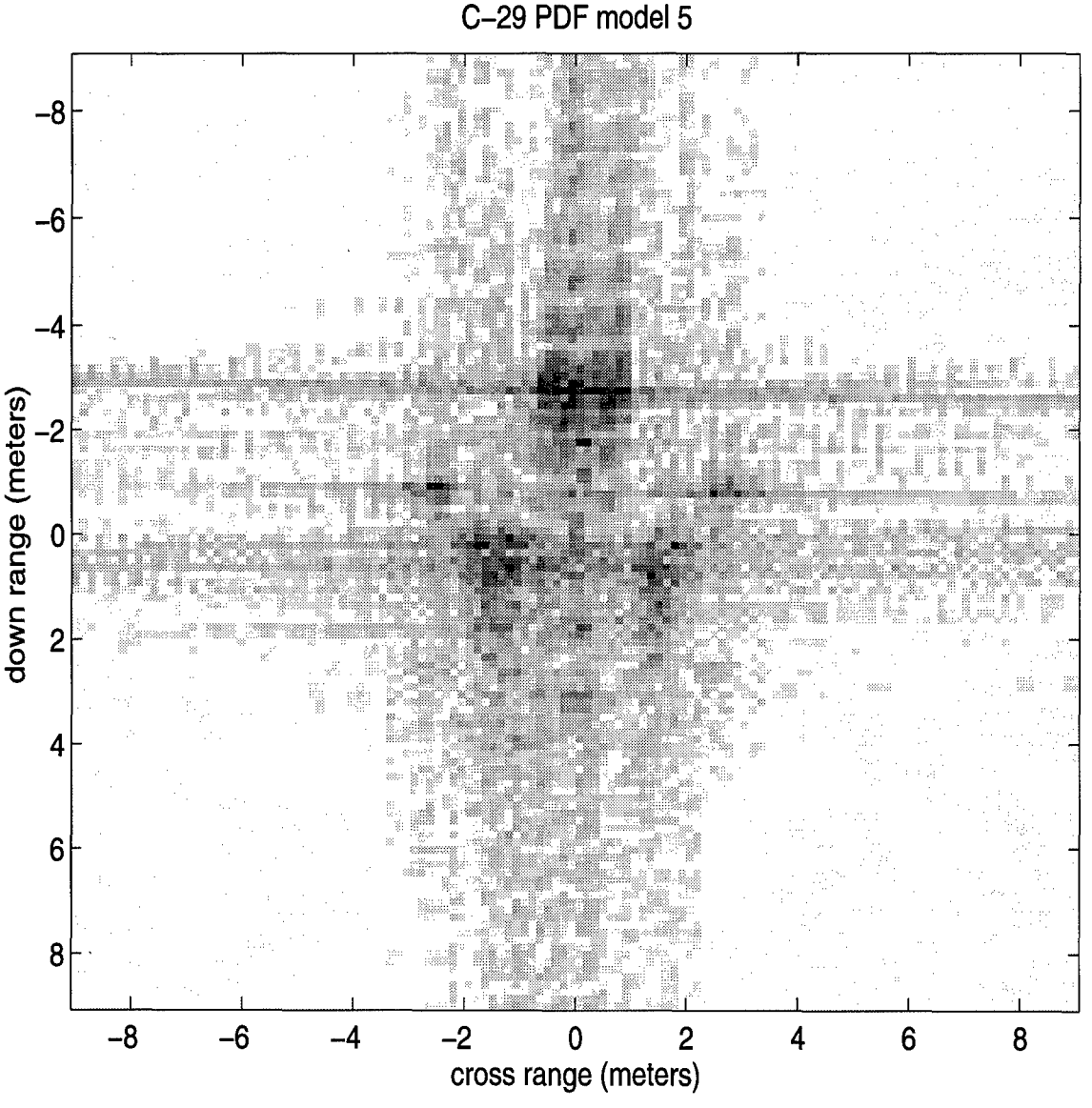


Figure 5. C-29 focused with the technique developed by Mensa [42]

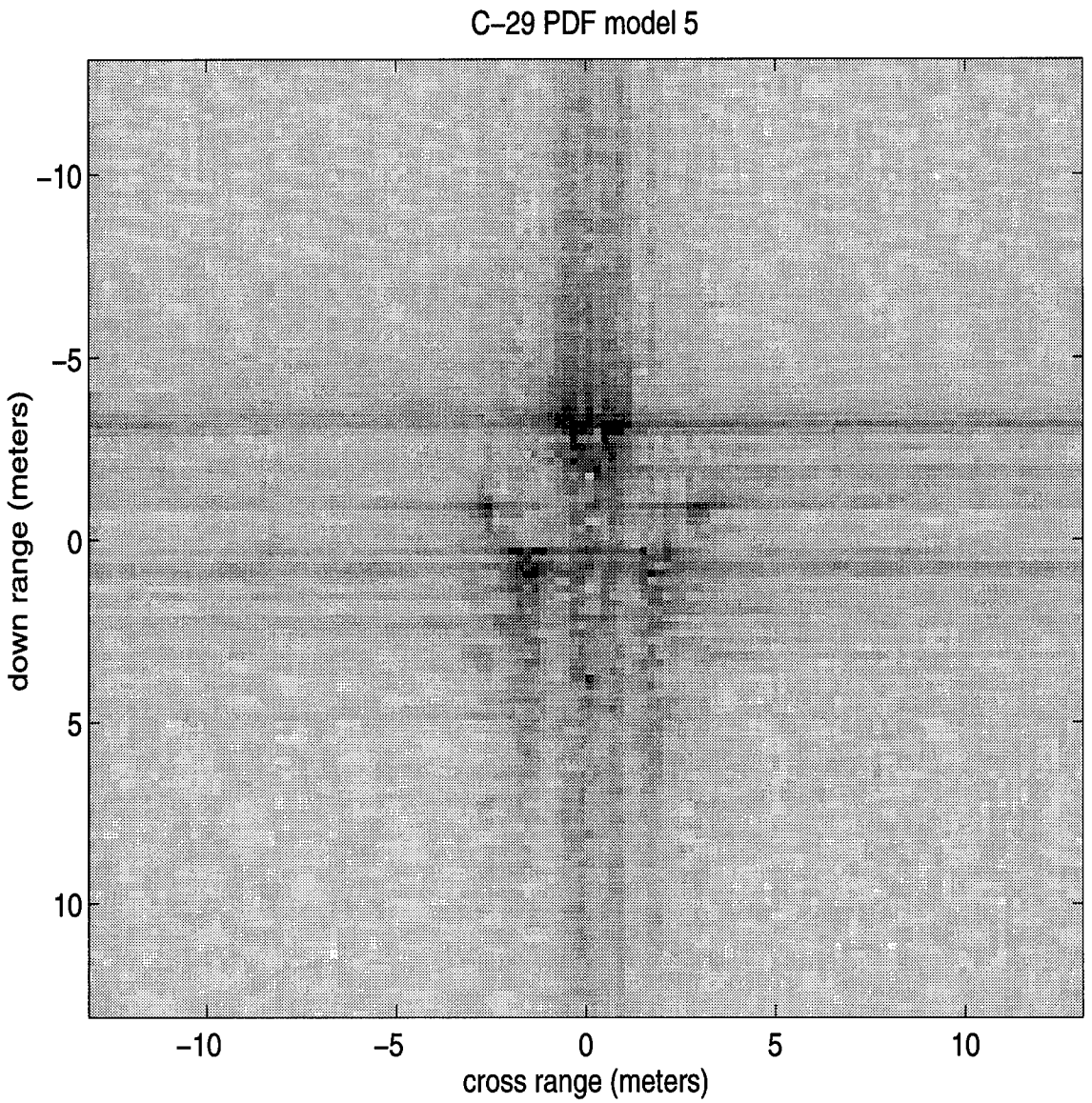


Figure 6. C-29 focused with scattering center focus (64 scattering centers)

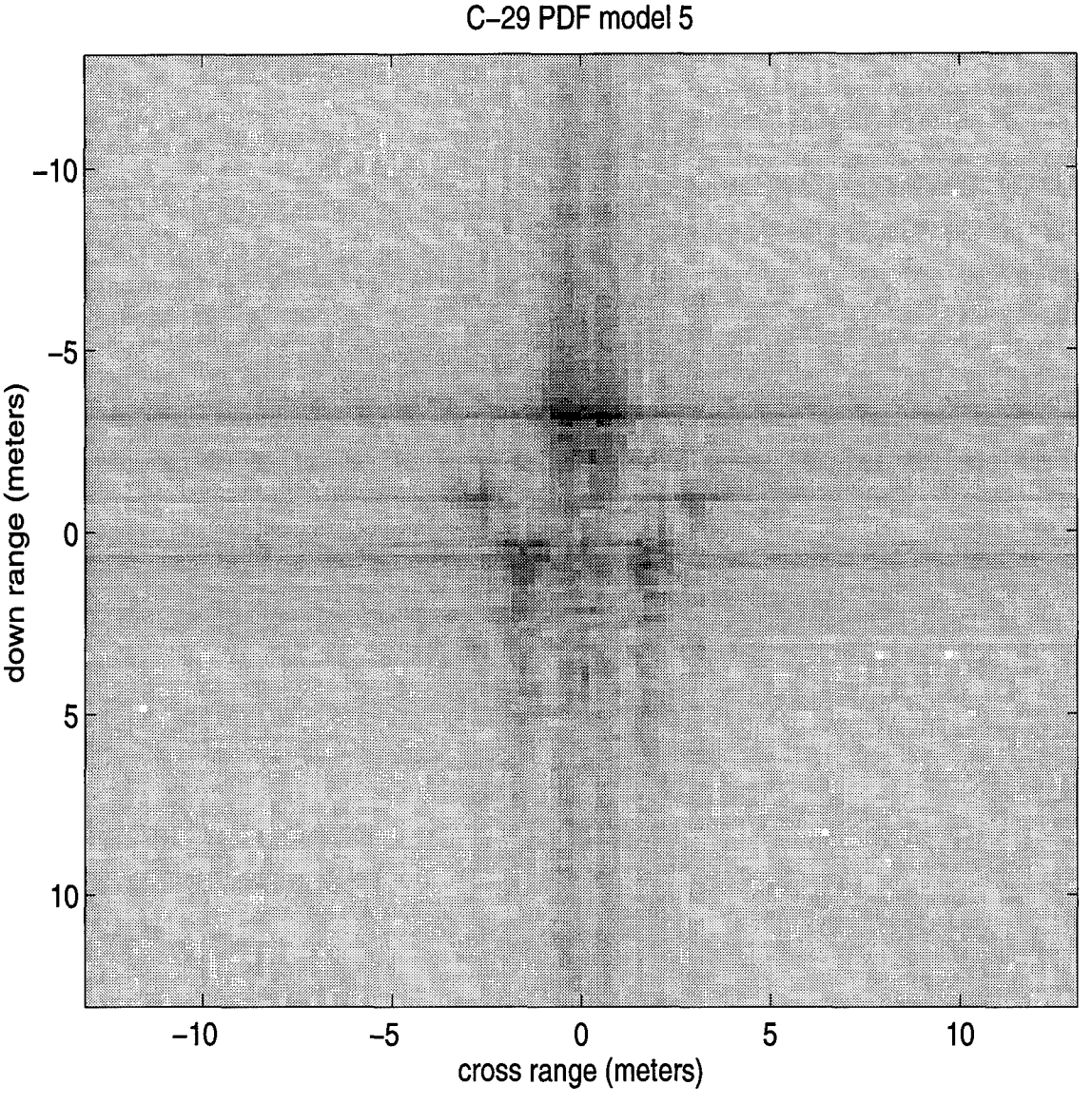


Figure 7. Unfocused image of C-29 (rectangular assumption)

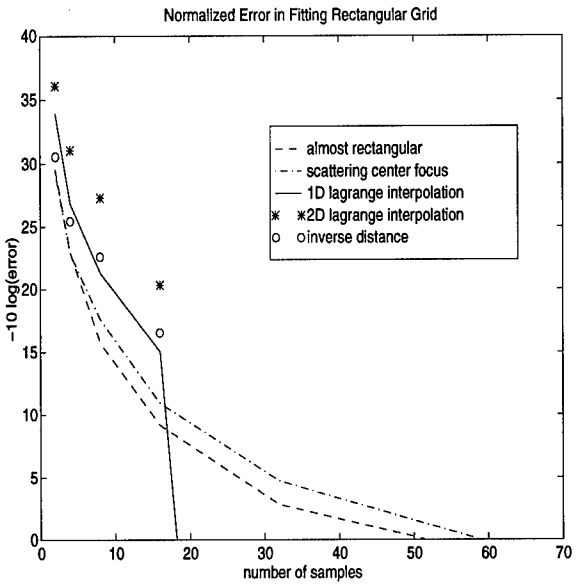


Figure 8. Interpolation error (Chamber Data)

lation methods. Mean Square Error (MSE) for phase history data generated from Equation 60 for two exponentials with parameters $(s, \lambda, \gamma) = \{(0.9e^{-j0.1\pi}, 0.97e^{-j0.2\pi}, e^{j0.2\pi}), (e^{j0.2\pi}, 1, 0.95e^{j0.4\pi})\}$ and white Gaussian noise in an $8 \times 8$ snapshot were calculated using Monte Carlo simulations each with 50 independent experiments. Several 2-D methods [10] [25] [38] [59] [2] were tested for accuracy without interpolation, with focusing, and with inverse imaging. These methods attain either the Cramér-Rao bound (CRB) for performing the estimation independently in each dimension (a multi-trial 1-D bound or the 1-D by 1-D) or the 2-D CRB [11] [52] [66]. Figures 9 through 12 show the Cramér-Rao bounds and resultant MSE for the estimated frequency of one exponential. Two solid lines are displayed; the upper is the 2-D Cramér-Rao bound and the lower is the multi-trial 1-D bound.

Degradation away from the Cramér-Rao bound is observed even for this small $8 \times 8$ snapshot. The 'almost rectangular' assumption fails quickly. When the sample grid is greater than $32 \times 32$, the pixel boundary condition is violated and no estimation technique performs well with this assumption. Each of the interpolation methods improves the performance of the algorithms in spatial frequency estimation (scatterer location). The performance of some estimation methods, however, is degraded and no longer attains the Cramér-Rao bound.

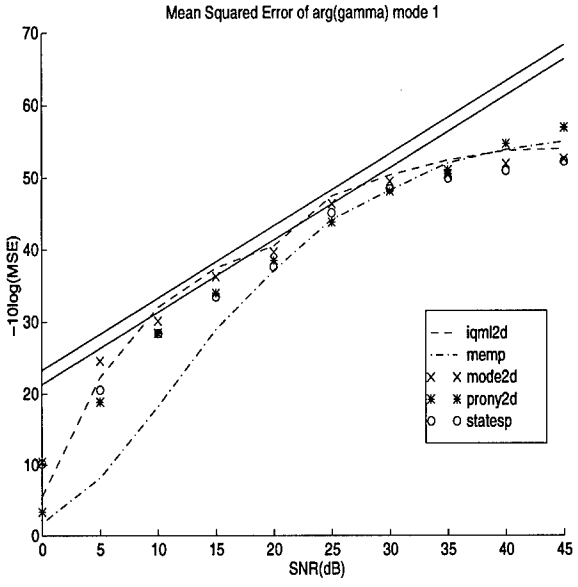


Figure 9. MSE plot for almost rectangular assumption. Techniques: 2-D iterative quadratic maximum likelihood (iqml2d), matrix enhancement matrix pencil (memp), 2-D method of direction estimation (mode2d), 2-D prony (prony2d), and state space (statesp) methods

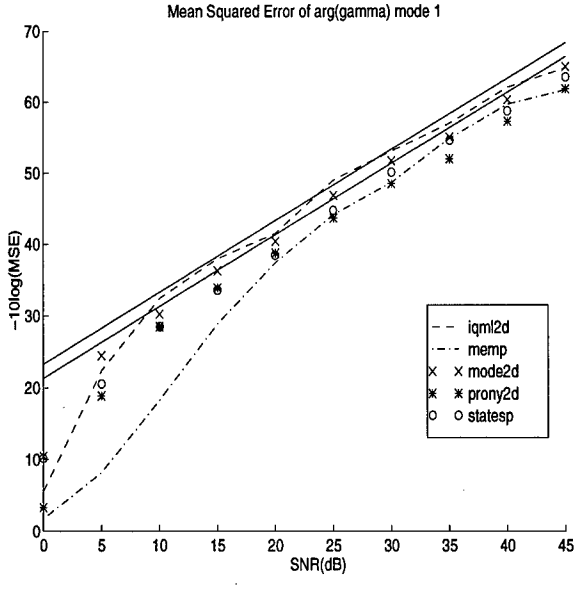


Figure 10. MSE plot for scattering center focus. Techniques: 2-D iterative quadratic maximum likelihood (iqml2d), matrix enhancement matrix pencil (memp), 2-D method of direction estimation (mode2d), 2-D prony (prony2d), and state space (statesp) methods

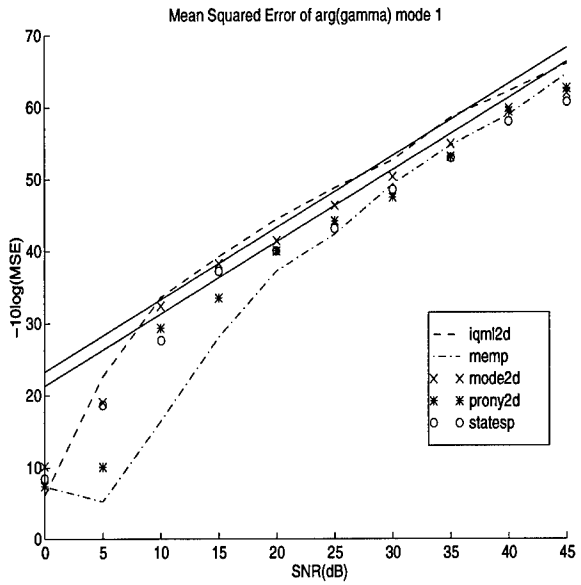


Figure 11. MSE plot for 1-D Lagrange interpolation. Techniques: 2-D iterative quadratic maximum likelihood (iqml2d), matrix enhancement matrix pencil (memp), 2-D method of direction estimation (mode2d), 2-D prony (prony2d), and state space (statesp) methods

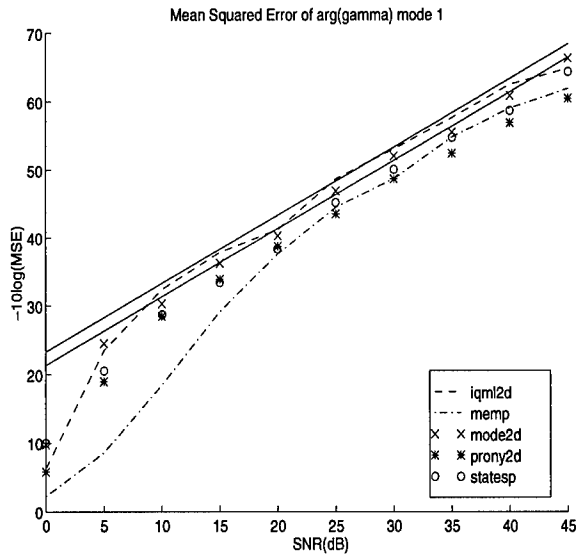


Figure 12. MSE plot for 2-D inverse distance interpolation. Techniques: 2-D iterative quadratic maximum likelihood (iqml2d), matrix enhancement matrix pencil (memp), 2-D method of direction estimation (mode2d), 2-D prony (prony2d), and state space (statesp) methods

The benefit of combining interpolation and estimation is easily seen. Not only will the interpolation method then be appropriate to the specific method; but the estimation technique could also better exploit the interpolation. The non-rectangular data grid could be used directly by the algorithm, perhaps to form a projection onto the signal subspace. Or the interpolation results such as from the Lagrange polynomial could be used instead of the interpolated data. The interpolation problem is not as significant for airborne data shown in Figure 3 where the $\mathbf{k} - \omega$ domain data is evenly sampled in the $k_u$ direction, however, it remains to be seen how flight path deviations and other real-world problems effect scatterer location.

# *V. Maximum Likelihood Estimation of 1-D Exponentials in Colored Noise*

## *5.1 Overview*

Parametric techniques to estimate complex exponential signals in noise play an important role in many signal processing applications such as direction of arrival estimation in array processing, high resolution spectral estimation, and radar data modeling. An overview of these diverse techniques is found in [66]. A key assumption in many of these techniques is that the noise spectrum is flat or white. When the noise is not white, more accurate estimation of the exponentials is possible. White noise techniques do not take advantage of this fact. Additionally, peaks in the noise spectrum may be incorrectly identified as exponentials. This work in this chapter does not rely on the white noise assumption. Instead, exponential estimation techniques are based on the assumption that the noise is unknown, allowing noise with spectral variation (colored noise) to be parameterized and estimated. Because of superior statistical performance, maximum likelihood (ML) estimators of the parameters are developed. Several approaches [5] [36] [77], [86] to find the maximum likelihood estimate of exponentials in colored noise have been developed that model the signal as random amplitude exponentials (the stochastic model). Most of these methods, however, require several instances of data to form an estimate of the data covariance matrix. Stochastic ML techniques are examined in the latter half of this chapter. This chapter is primarily concerned with single data instances. In this case, the exponential frequencies and amplitudes can be considered unknown constants (the deterministic model). An approximation to the ML estimator for the parameters of a colored noise deterministic model was developed in [29]. Section 4.3 shows how the method in [29] approximates the true ML estimator derived in this chapter. This chapter considers the estimation of the most general case of exponential, the damped exponential. The methods developed can be easily extended to undamped exponentials with appropriate constraints. The methods are also easily extended for multiple instances of data.

This chapter employs the traditional approach to finding the ML estimators of the exponential parameters in the deterministic model. The log-likelihood function of the data and parameters is determined by taking the logarithm of the probability density function of the observed data. When a parameter can be decoupled and estimated separately, the derivative of the log-likelihood function with respect to the parameter is determined. The derivative is then set equal to zero in order to find the value of the parameter that maximizes the likelihood function, i.e., the ML estimator of

that parameter. ML estimators for the exponential amplitudes and the noise variance are found in this manner. When these estimates are substituted back into the log-likelihood function, the resulting compressed log-likelihood function is a nonlinear function of the exponential frequencies and the noise parameters. Computationally efficient methods to find the parameters that maximize this compressed log-likelihood function are developed by extending the white noise IQML method [6] [33] [66]. Unknown colored noise techniques are developed for the cases of stationary noise, autoregressive (AR) noise (white noise driving an all-pole filter), and autoregressive moving-average (ARMA) noise (white noise driving a filter with poles and zeros).

This chapter is organized as follows. Section 5.2 reviews the damped exponential model and develops the maximum likelihood estimate of the parameters of this model, develops methods for efficient computation of the deterministic ML estimates for several colored noise models, and shows the statistical performance of these new methods with respect to the Cramér-Rao estimation bound. Section 5.3 develops efficient methods of solving the stochastic maximum likelihood colored noise problem and shows the statistical performance of these methods.

## *5.2 1-D Deterministic Maximum Likelihood*

*5.2.1 1-D Damped Exponential Model.* This section describes the damped exponential model with colored noise. Then deterministic maximum likelihood is applied to estimate the parameters of the model. The estimation of the damped exponential model for a single data instance is examined. This case is easily modified for the multiple instance case, or the constrained to the unit circle case by techniques described in [6] and [34].

The complex valued data samples, $y[n]$ for $n = 0 \ldots N-1$, of the damped exponential model are given by

$$y[n] = \sum_{i=1}^{p} s_i \lambda_i^n + w[n], \quad n = 0 \ldots N-1, \tag{66}$$

where $p$ is the number of exponentials, $s_i$ is the amplitude of the $i^{th}$ exponential, $\lambda_i^n = e^{-\alpha_i n + j\beta_i n}$ is the $i^{th}$ damped exponential with damping factor $\alpha_i$ and exponential frequency $\beta_i$, and $w[n]$ is a zero-mean, stationary Gaussian noise sequence. The exponential part of the signal is considered deterministic with unknown parameters, $\{\lambda = [\ \lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_p\ ]^T, \mathbf{s} = [\ s_1 \quad s_2 \quad \cdots \quad s_p\ ]^T\}$. The randomness of the data is due entirely to the normally distributed noise sequence, $\mathbf{w}(\sigma^2, \eta) \sim$

$\mathrm{N}[0, \sigma^2 R_{\mathbf{ww}}(\eta)]$, where $\sigma^2$ is the noise variance, and $R_{\mathbf{ww}}(\eta)$ is the noise covariance matrix with $q$ unknown parameters $\{\eta = [\ \eta_1 \quad \eta_2 \quad \cdots \quad \eta_q\ ]^T\}$. When the data samples are arranged in a column vector, Equation 66 is written with respect to the model parameters $\theta = \{\lambda, \mathbf{s}, \sigma^2, \eta\}$ as

$$\mathbf{y}(\theta) = G(\lambda)\mathbf{s} + \mathbf{w}(\sigma^2, \eta), \tag{67}$$

where $G(\lambda)$ is an $N \times p$ matrix with a Vandermonde column of each of the damped exponentials such that

$$\begin{aligned} G(\lambda) &= [\ \mathbf{g}_1(\lambda_1) \quad \mathbf{g}_2(\lambda_2) \quad \cdots \quad \mathbf{g}_p(\lambda_p)\ ], \\ \mathbf{g}_i(\lambda_i) &= [\ 1 \quad \lambda_i \quad \lambda_i^2 \quad \cdots \quad \lambda_i^N\ ]^T, \end{aligned} \tag{68}$$

and $\mathbf{s}$ is a column vector of the exponential amplitudes. Additionally, $\mathbf{w}$ is complex circularly symmetric such that $\mathrm{E}\{\mathbf{w}^*\mathbf{w}\} = \sigma^2 R_{\mathbf{ww}}(\eta)$ and $\mathrm{E}\{\mathbf{w}^T\mathbf{w}\} = 0$, where $(\cdot)^T$ indicates transpose and $(\cdot)^*$ indicates complex conjugate transpose. Thus the data $\mathbf{y}(\theta)$ is normally distributed as $\mathrm{N}[G(\lambda)\mathbf{s}, \sigma^2 R_{\mathbf{ww}}(\eta)]$.

*5.2.2 Maximum Likelihood Estimators.* The maximum likelihood estimate of the parameters of the damped exponential and colored noise model, $\theta = \{\lambda, \mathbf{s}, \sigma^2, \eta\}$, are those values of the parameters that maximize the probability density function for the observed data sample $\mathbf{y}$. For simplicity, the explicit dependence of $G$ and $R_{\mathbf{ww}}$ on the model parameters $\lambda$ and $\eta$, respectively, is dropped from the notation. The probability density to be maximized is the circular complex Gaussian density,

$$f(\mathbf{y}) = \frac{1}{|\pi\sigma^2 R_{\mathbf{ww}}|} e^{-\frac{1}{\sigma^2}(\mathbf{y}-G\mathbf{s})^* R_{\mathbf{ww}}^{-1}(\mathbf{y}-G\mathbf{s})}. \tag{69}$$

The logarithm is a monotonic function, thus the logarithm of the probability density is equivalently maximized with respect to parameters in $\lambda, \mathbf{s}, \sigma^2$ and $\eta$. This log-likelihood function for the density in Equation 69 is

$$\mathcal{L}(G, \mathbf{s}, \sigma^2, R_{\mathbf{ww}}; \mathbf{y}) = -N \ln \pi\sigma^2 - \ln |R_{\mathbf{ww}}| - \frac{1}{\sigma^2}(\mathbf{y} - G\mathbf{s})^* R_{\mathbf{ww}}^{-1}(\mathbf{y} - G\mathbf{s}). \tag{70}$$

The values of the parameters which maximize this function are the maximum likelihood estimates of those parameters. The derivatives of the log-likelihood function with respect to $\sigma^2$ and $\mathbf{s}$ are

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = -\frac{N}{\sigma^2} + \frac{1}{\sigma^4}(\mathbf{y} - G\mathbf{s})^* R_{\mathbf{ww}}^{-1}(\mathbf{y} - G\mathbf{s}), \tag{71}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{s}^*} = -G^* R_{\mathbf{ww}}^{-1} G\mathbf{s} + G^* R_{\mathbf{ww}}^{-1}\mathbf{y} \tag{72}$$

where the derivatives with respect to complex vectors are defined in [74]. Setting these derivatives equal to zero gives the maximum likelihood estimators of $\sigma^2$ and $\mathbf{s}$,

$$\hat{\sigma}^2 = \frac{1}{N}(\mathbf{y} - G\mathbf{s})^* R_{\mathbf{ww}}^{-1}(\mathbf{y} - G\mathbf{s}), \tag{73}$$

$$\hat{\mathbf{s}} = (G^* R_{\mathbf{ww}}^{-1} G)^{-1} G^* R_{\mathbf{ww}}^{-1}\mathbf{y}. \tag{74}$$

A compressed log-likelihood function can then be constructed by substituting the ML estimates of $\sigma^2$ and $\mathbf{s}$ into the log-likelihood function,

$$\mathcal{L}(G, R_{\mathbf{ww}}; \mathbf{y}) \equiv \mathcal{L}(G, \mathbf{s}, \sigma^2 R_{\mathbf{ww}}; \mathbf{y}) \mid_{\sigma^2 = \hat{\sigma}^2, \mathbf{s} = \hat{\mathbf{s}}} \tag{75}$$

$$= -N \ln \frac{\pi}{N} \mathbf{y}^* (I - G(G^* R_{\mathbf{ww}}^{-1} G)^{-1} G^* R_{\mathbf{ww}}^{-1})^* R_{\mathbf{ww}}^{-1} (I - G(G^* R_{\mathbf{ww}}^{-1} G)^{-1} G^* R_{\mathbf{ww}}^{-1})\mathbf{y} - \ln |R_{\mathbf{ww}}| - N.$$

The remaining parameters are found by minimizing the negative of the compressed log-likelihood function. Combining terms and ignoring constant terms the minimization is

$$\min_{\lambda, \eta} \ln |\mathbf{y}^* (I - G(G^* R_{\mathbf{ww}}^{-1} G)^{-1} G^* R_{\mathbf{ww}}^{-1})^* R_{\mathbf{ww}}^{-1} (I - G(G^* R_{\mathbf{ww}}^{-1} G)^{-1} G^* R_{\mathbf{ww}}^{-1})\mathbf{y} R_{\mathbf{ww}}|. \tag{76}$$

This expression is quickly simplified when we define a projection onto the signal space

$$P_{L,G} \equiv L^{-*} G (G^* R_{\mathbf{ww}}^{-1} G)^{-1} G^* L^{-1} \tag{77}$$

where the positive definite covariance matrix is decomposed as $R_{\mathbf{ww}} = L^* L$, and $L^{-*} = (L^*)^{-1}$. The minimization is then

$$\min_{\lambda, \eta} \ln |\mathbf{y}^* (I - L^* P_{L,G} L^{-*})^* R_{\mathbf{ww}}^{-1} (I - L^* P_{L,G} L^{-*})\mathbf{y} R_{\mathbf{ww}}|$$

$$
\begin{aligned}
&= \min_{\lambda,\eta} \ln |\mathbf{y}^*(L^*L^{-*} - L^*P_{L,G}L^{-*})^* R_{\mathbf{ww}}^{-1}(L^*L^{-*} - L^*P_{L,G}L^{-*})\mathbf{y}R_{\mathbf{ww}}| \\
&= \min_{\lambda,\eta} \ln |\mathbf{y}^*L^{-1}(I - P_{L,G})^* L R_{\mathbf{ww}}^{-1} L^*(I - P_{L,G})L^{-*}\mathbf{y}R_{\mathbf{ww}}| \\
&= \min_{\lambda,\eta} \ln |\mathbf{y}^*L^{-1}(I - P_{L,G})L^{-*}\mathbf{y}R_{\mathbf{ww}}| \qquad (78)
\end{aligned}
$$

since the projection $(I - P_{L,G})$ is idempotent and Hermitian symmetric. Next, we define a projection onto the space orthogonal to the signal subspace [9],

$$P_{L,A} = LA(A^*R_{\mathbf{ww}}A)^{-1}A^*L^*, \qquad (79)$$

where $A^*G = 0$. The $N-p \times N$ matrix $A$ is formed from the coefficients ($\mathbf{a} = [\ a_1 \quad a_2 \quad \cdots \quad a_p\ ]^T$) of a polynomial,

$$a(\lambda_i) = a_0 + a_1\lambda_i^{-1} + a_2\lambda_i^{-2} + \ldots + a_p\lambda_i^{-p} = 0 \quad (i = 1 \ldots p), \qquad (80)$$

that annihilates all of the exponential columns in $G$

$$A^* = \begin{bmatrix} a_p & a_{p-1} & \cdots & a_0 & & 0 \\ & \ddots & \ddots & & \ddots & \\ 0 & & a_p & a_{p-1} & \cdots & a_0 \end{bmatrix}. \qquad (81)$$

The $\lambda_i$ $i = 1 \ldots p$ are simply the roots of the polynomial $a(\lambda)$. Since $G$ is rank $p$, $A$ is rank $N - p$, and $G$ and $A$ are orthogonal, the projections $P_{L,G}$ and $P_{L,A}$ form a resolution of the identity where

$$P_{L,G} + P_{L,A} = I_{N \times N}. \qquad (82)$$

Thus, the minimization in Equation 78 in terms of the space orthogonal to the signal space,

$$
\begin{aligned}
&\min_{a,\eta} \ln |\mathbf{y}^*L^{-1}P_{L,A}L^{-*}\mathbf{y}R_{\mathbf{ww}}| \\
&= \min_{a,\eta} \ln |\mathbf{y}^*A(A^*R_{\mathbf{ww}}A)^{-1}A^*\mathbf{y}R_{\mathbf{ww}}| \\
&= \min_{a,\eta} \ln |\mathbf{y}^*A(A^*R_{\mathbf{ww}}A)^{-1}A^*\mathbf{y}I| + \ln |R_{\mathbf{ww}}|, \qquad (83)
\end{aligned}
$$

will give the ML estimates of the parameters $\mathbf{a}$, the annihilator of the exponentials, and $\eta$, the noise parameters. This is a nonlinear, multidimensional optimization problem which could be solved by

many well known methods. Notice, however, that the derivative with respect to $\mathbf{a}$ of the terms to be minimized in Equation 83 does not include the second term. The ML estimator of the exponentials does not depend on this noise related term. It is also suggested in [30] when estimating noise that the second term in the equation simply constrains the potential solutions (noise poles) away from the unit circle.

Performing the minimization in Equation 83 separately with respect to $\mathbf{a}$, the minimization gives

$$\min_{\mathbf{a}} \mathbf{y}^* A(A^* R_{\mathbf{ww}} A)^{-1} A^* \mathbf{y}. \tag{84}$$

Further, the inverse in Equation 84 is decomposed as $(A^* R_{\mathbf{ww}} A)^{-1} = (A^* L^* L A)^{-1} = (LA)^+ (A^* L^*)^+$, where $(\cdot)^+$ indicates the Moore-Penrose pseudoinverse, by using the pseudoinverse formula $(A^* L^*)^+ = LA(A^* L^* LA)^{-1}$ and $(LA)^+ = (A^* L^* LA)^{-1} A^* L^*$. The minimization is then stated compactly and efficiently computed as

$$\begin{aligned} & \min_{\mathbf{a}} \mathbf{y}^* A(LA)^+ (A^* L^*)^+ A^* \mathbf{y} \\ = \; & \min_{\mathbf{a}} \left\| (A^* L^*)^+ A^* \mathbf{y} \right\|_F^2 \end{aligned} \tag{85}$$

where $(LA)^*(LA)$ is the Cholesky decomposition of $A^* R_{\mathbf{ww}} A$ and $\|\cdot\|_F^2$ is the Frobenius norm. The following section shows how $R_{\mathbf{ww}}$ is attained and how the minimization in Equation 85 can be efficiently computed.

*5.2.3 Conditional Estimation of the Exponentials.* This section introduces a new method for estimation of exponentials in unknown colored noise. With this method an efficient white noise technique is extended to unknown colored noise for several colored noise models. Several new algorithms are then developed to perform this estimation.

The asymptotic properties of maximum likelihood are retained when Equation 84 is optimized with respect to $\mathbf{a}$ by using an asymptotically unbiased, consistent estimate of $R_{\mathbf{ww}}$. Given a consistent estimate of $R_{\mathbf{ww}}$, $\hat{R}_{\mathbf{ww}}$, the minimization in Equation 83 is asymptotically equivalent to

$$\min_{\mathbf{a}} \mathbf{y}^* A(A^* \hat{R}_{\mathbf{ww}} A)^{-1} A^* \mathbf{y}. \tag{86}$$

As is the case with other ML estimates, in addition to asymptotic efficiency, Equation 86 produces good estimates for short data records as well. For the case of white noise, IQML [6] [33] [66] is a computationally efficient, iterative method to find the parameters $\mathbf{a}$ in Equation 86. In IQML, the $A^*\mathbf{y}$ term is restated as $Y\mathbf{a}$ where

$$A^*\mathbf{y} = \begin{bmatrix} a_p & a_{p-1} & \cdots & a_0 & & 0 \\ & \ddots & \ddots & & \ddots & \\ 0 & & a_p & a_{p-1} & \cdots & a_0 \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} = \begin{bmatrix} y_p & y_{p-1} & \cdots & y_0 \\ y_{p+1} & y_p & \cdots & y_1 \\ \vdots & & \ddots & \vdots \\ & & & y_p \\ \vdots & & & \vdots \\ y_{N-1} & \cdots & & y_{N-p-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = Y\mathbf{a}. \tag{87}$$

With Equation 87, the maximum likelihood minimization in Equation 86 becomes

$$\min_{\mathbf{a}} \mathbf{a}^* Y^* (A^* \hat{R}_{\mathbf{ww}} A)^{-1} Y \mathbf{a}. \tag{88}$$

In IQML, $A$ is constructed from the elements of $\mathbf{a}$ in a previous iteration. The quadratic form in Equation 88 is minimized by any of a number of well-known methods. The constraint $\|\mathbf{a}\|_F = 1$ guarantees a unique solution and the minimizing $\mathbf{a}$ is then the eigenvector associated with the minimum eigenvalue of $Y^*(A^* \hat{R}_{\mathbf{ww}} A)^{-1} Y$.

The basic iteration used in all the algorithms developed in this section is summarized at the $k^{th}$ iteration as

$$\min_{\mathbf{a}_k} \mathbf{a}_k^* Y^* (A_{k-1}^* \hat{R}_{\mathbf{ww}} A_{k-1})^{-1} Y \mathbf{a}_k. \tag{89}$$

The kernel $Y^*(A_{k-1}^* \hat{R}_{\mathbf{ww}} A_{k-1})^{-1} Y$ is quickly calculated using the Cholesky decomposition of $A_{k-1}^* \hat{R}_{\mathbf{ww}} A_{k-1}$ in a colored noise implementation of [23]. To initialize the algorithm, $A$ is set equal to the identity matrix $I$. Thus, the first iteration whitens the data with the covariance $\hat{R}_{\mathbf{ww}}^{-1}$ before estimating the exponential frequencies.

To demonstrate the achievable performance of the algorithms developed in this section the known colored noise version of IQML where $\hat{R}_{\mathbf{ww}} = R_{\mathbf{ww}}$ in Equation 89 was implemented and is called Colored noise IQML (CIQML).

*5.2.4 Estimation in Unknown Colored Noise.* Imposing some assumptions in order to parameterize the noise, allows development of consistent estimators of the parameters. In all the following cases, the noise is assumed ergodic. Thus, the autocorrelation sequence of the noise is consistently estimated [51] as

$$\hat{r}_{ww}[m] = \frac{1}{N}\sum_{n=0}^{N-m} w[n]w[n+m] \qquad m = -(N-1)\ldots 0\ldots N-1. \tag{90}$$

The problem posed by Equation 89 is how to initiate the iteration with a good estimate of $\mathbf{a}$ and $R_{\mathbf{ww}}$. Better estimates are attained when the signal and noise are estimated simultaneously. This, however, is not always possible.

*5.2.4.1 Unknown Colored Noise IQML (UCIQML).* A common assumption made is that the noise is stationary. This is the case solved in [77] for the stochastic model. The matrix $R_{\mathbf{ww}}$ is then Toeplitz, Hermitian symmetric with $N$ parameters. These $N$ parameters of the autocorrelation function are based on $N$ real-valued of the power spectral density (PSD). Then, together with the undamped exponential signal, $4p+N+1$ real-valued parameters are to be estimated. For the single data instance, this problem is well-posed since $N$ complex-valued samples of data are available. For this case, the signal and noise are estimated separately as detailed below.

White noise exponential estimators are consistent in colored noise [21]. That is asymptotically, the residual noise, the difference between the estimate and the observed data, can not be used to improve the exponential estimate. It is, however, the good performance of ML estimators such as IQML for short data records, that allows us to estimate the signal and subtract it from the data to estimate the noise sequence. From Equation 67 the noise sequence is consistently estimated (see Appendix B) as

$$\mathbf{w} = \mathbf{y} - G\mathbf{s}, \tag{91}$$

where $G$ contains Vandermonde columns of exponentials estimated by IQML and $\mathbf{s}$ is estimated from Equation 74 with $R_{\mathbf{ww}} = I$ as

$$\mathbf{s} = (G^*G)^{-1}G^*\mathbf{y} = G^+\mathbf{y}, \tag{92}$$

where $(\cdot)^+$ indicates the pseudoinverse. $R_{\mathbf{ww}}$ is simply the Toeplitz covariance matrix

$$R_{\mathbf{ww}} = \begin{bmatrix} \hat{r}_{\mathbf{ww}}[0] & \hat{r}^*_{\mathbf{ww}}[1] & \cdots & \hat{r}^*_{\mathbf{ww}}[N-1] \\ \hat{r}_{\mathbf{ww}}[1] & \ddots & \ddots & \\ \vdots & \ddots & & \hat{r}^*_{\mathbf{ww}}[1] \\ \hat{r}_{\mathbf{ww}}[N-1] & & \hat{r}_{\mathbf{ww}}[1] & \hat{r}_{\mathbf{ww}}[0] \end{bmatrix} \tag{93}$$

with elements consisting of the autocorrelation sequence of $\mathbf{w}$, $\hat{r}_{\mathbf{ww}}[m]$ for $m = -(N-1)\ldots N-1$. With the estimate of $\mathbf{a}$ from IQML, and $L$ from the Cholesky decomposition of $R_{\mathbf{ww}}$, the iteration in Equation 89 can be computed. Additionally, since this iteration improves the estimate of $\mathbf{a}$, and consequently the estimate of $\mathbf{s}$, $\mathbf{s}$ is reestimated at each iteration by Equation 74 and $\mathbf{w}$ is reestimated by Equation 91. For a speedy implementation, the autocorrelation sequence, $\hat{r}_{\mathbf{ww}}[m]$ for $m = -(N-1)\ldots N-1$, is quickly computed with the Fast Fourier Transform (FFT) and estimates are quickly computed with the pseudoinverse computed by the QR decomposition. For example Equation 74 becomes

$$\hat{\mathbf{s}} = ((L^*)^+ G)^+ (L^*)^+ \mathbf{y}. \tag{94}$$

Just as in IQML of the iteration in Equation 89 is not guaranteed to converge, however, after a small number of iterations (less than ten) estimates are significantly improved. One key element of sustaining the iteration in Equation 89 is maintaining the positive definiteness of the kernel $(A^* R_{\mathbf{ww}} A)^{-1}$. Thus, only positive definite estimates of the noise covariance, $R_{\mathbf{ww}}$, are used in the methods developed in this dissertation.

*5.2.4.2 The UCIQML Algorithm.* The proceeding algorithm will be called Unknown Colored noise IQML (UCIQML) and is summarized as follows:

Step 1. *Estimate the signal* $G\mathbf{s}$ *with IQML.* Set $A_0 = I$ and $R_{\mathbf{ww}} = I$ in Equation 89. Iterate several times to attain the roots of $\mathbf{a}$, $\lambda_i$ for $i = 1\ldots p$. Form the matrix $G$ from the $\lambda_i$. Then estimate $\mathbf{s}$ as

$$\mathbf{s} = G^+ \mathbf{y}. \tag{95}$$

Step 2. *Estimate the noise sequence* $\mathbf{w}$ as

$$\mathbf{w} = \mathbf{y} - G\mathbf{s}. \tag{96}$$

Step 3. *Estimate the autocorrelation sequence of the noise* $r_{\mathbf{ww}}[m]$ for $m = -(N-1)\ldots N-1$. Form $R_{\mathbf{ww}}$ and its Cholesky decomposition $R_{\mathbf{ww}} = L^*L$.

Step 4. *Estimate new* $\lambda_i$ for $i = 1\ldots p$. Form $A_{k-1}$ from the last estimate of $\mathbf{a}$, $\mathbf{a}_{k-1}$, and iterate once

$$\min_{\mathbf{a}_k} \left\| (A^*_{k-1}L^*)^+ Y \mathbf{a}_k \right\|^2_F . \tag{97}$$

Step 5. *Re-estimate* $\mathbf{s}$ as

$$\hat{\mathbf{s}} = ((L^*)^+ G)^+ (L^*)^+ \mathbf{y}. \tag{98}$$

Step 6. *Repeat* Step 2 - 5 for several iterations, or until estimates do not significantly change.

*5.2.4.3 Other Noise Models.* In addition to stationarity, the noise may be further parameterized as the stationary output of a filter driven by a white noise sequence. Since the available data is limited, an appropriate low-order model must be chosen to limit the number of parameters to be estimated. The principle of parsimony [7] suggests that to better estimate the signal, the noise be modeled with as few parameters as possible.

Two well-known models that provide adjustable low-order noise models are the autoregressive (AR) and autoregressive moving-average (ARMA) filter models, with Power Spectral Densities (PSDs) being rational functions of two polynomials, $C(z)/B(z)$. In the next sections, specific solutions for AR and ARMA noise processes are explored. It is emphasized that the basic colored noise technique exploited in this research is very general and any appropriate low-order noise model may be used to estimate $R_{\mathbf{ww}}$ for Equation 89.

*5.2.4.4 Autoregressive Unknown Colored Noise IQML (ARUCIQML).* The following solution for the unknown colored noise problem assumes an AR model for the noise (i.e. the noise contains only poles and no zeros; the single filter coefficient of the numerator polynomial is one). If the noise sequence is the stationary output of a $q^{th}$ order Infinite Impulse Response (IIR)

filter, then the noise satisfies an autoregressive (AR) difference equation

$$b_0 w[n] = -b_1 w[n-1] \ \ldots - b_q w[n-q] + \epsilon[n] \tag{99}$$

where $\epsilon[n]$ is a white noise sequence. The AR PSD is

$$P_{ww}(e^{j\omega}) = \frac{\sigma^2}{\left|\sum_{n=0}^{q} b_n e^{-jn\omega}\right|^2}. \tag{100}$$

As stated earlier, the first problem posed by the minimization in Equation 89 is how to attain a good initial estimate of $R_{\mathbf{ww}}$ and $A$ so that the iteration will converge to the true $A$. It is well known that either AR noise poles or exponential frequencies may be estimated with the covariance method [29]. Additionally, IQML provides consistent ML estimates of both exponentials and AR processes [66] (see Appendix B). The only difference in modeling an AR noise sequence and exponentials is the input to the AR filter that produces them, white noise or an impulse. Figure 13 shows how the model of damped exponentials in colored noise includes both these cases. If we assume that the colored noise can be described with an autoregressive (AR), or all-pole filter model with known model order $q$, then the exponentials and noise poles may be estimated simultaneously by estimating an order $p+q$ AR polynomial with the covariance method or the IQML method.

*5.2.4.5 Approximating the Maximum Likelihood Optimization.* To see why this is true, notice that Equation 84 may be generalized as

$$\min_{\mathbf{a}} \mathbf{y}^* A W A^* \mathbf{y}, \tag{101}$$

where $W$ is positive definite. A generalization to any positive definite matrix $W$ is possible because the solution space for the minimization problem is defined by $A^*G = 0$ (the null space $A^*$). The matrix $AWA^*$ has the same null space. Next, decompose the inverse covariance matrix as in [30] and use the fact that $R_{\mathbf{ww}}^{-1}$ is centro-symmetric (Hermitian symmetric and symmetric about the anti-diagonal) [45] to produce

$$R_{\mathbf{ww}}^{-1} = BP^{-1}B^* = (BP^{-1/2})(P^{-1/2}B^*) = L^{-1}L^{-*}, \tag{102}$$

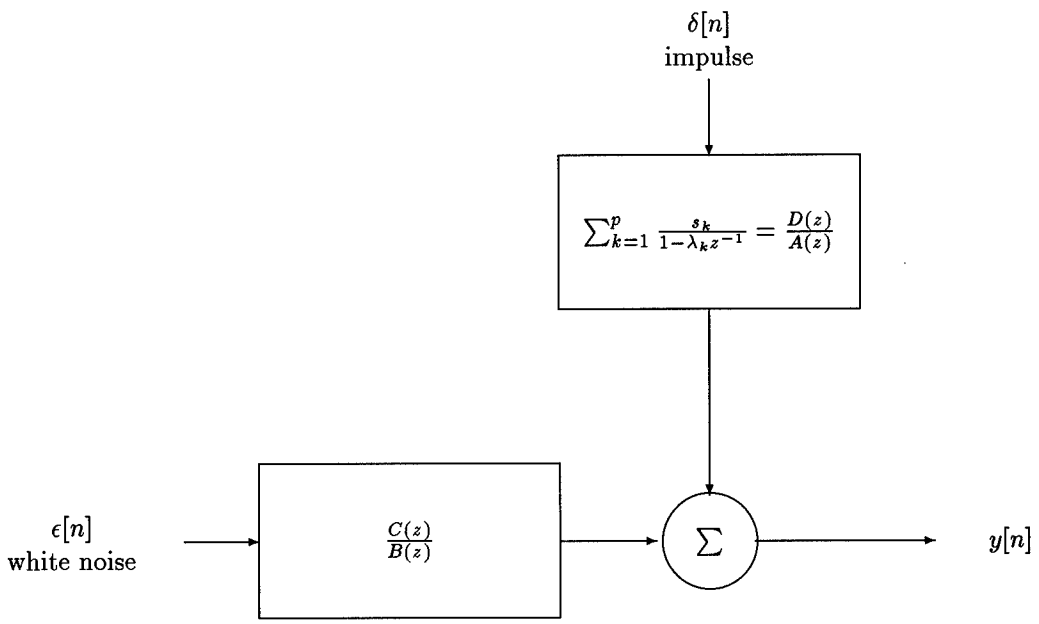


Figure 13. Model for exponentials in ARMA colored noise

where $B^*$ is the $N-q \times N$ matrix

$$B^* = \left[\begin{array}{ccccccc} & & & & & & \\ 1 & b_1 & & \cdots & & b_q & \mathbf{0} \\ & \ddots & \ddots & & & \ddots & \\ & & 1 & b_1 & \cdots & & b_q \\ \hline & & & 1 & b_1^{q-1} & \cdots & b_{q-1}^{q-1} \\ & & & & \ddots & & \vdots \\ & & & & & \ddots & b_1^1 \\ \mathbf{0} & & & & & & 1 \end{array}\right] = \left[\frac{K^*}{X}\right]. \tag{103}$$

The $b_i$ are the AR polynomial coefficients, $\mathbf{b} = [\ 1 \quad b_1 \quad \cdots \quad b_q\ ]$, the $b_i^k$ for $i = 1 \ldots k$ are the $k^{th}$ order AR coefficients from the Levinson recursion [30] and

$$P = \text{diag}([\ 1_1 \quad \cdots \quad 1_{N-q} \quad \rho_q \quad \cdots \quad \rho_0\ ])$$

where the $\rho_k$ for $k = 1 \ldots q$ are the prediction errors calculated at the $k^{th}$ iteration of the Levinson algorithm. The following approximations of Equation 84 asymptotically produce the same result and approximate the maximum likelihood solution,

$$\begin{aligned}
& \min_A \mathbf{y}^* A (A^* R_{\mathbf{ww}} A)^{-1} A^* \mathbf{y} \\
= \; & \min_A \mathbf{y}^* A (A^* (B P^{-1} B^*)^{-1} A)^{-1} A^* \mathbf{y} && (104) \\
\simeq \; & \min_A \mathbf{y}^* A (A^* (K K^*)^{+} A)^{-1} A^* \mathbf{y} \quad (q \leq p) && (105) \\
\simeq \; & \min_A \mathbf{y}^* A \tilde{K} (\tilde{A}^* \tilde{A})^{-1} \tilde{K}^* A^* \mathbf{y} && (106) \\
\simeq \; & \min_A \mathbf{y}^* A \tilde{K} \tilde{K}^* A^* \mathbf{y}. && (107)
\end{aligned}$$

In Equation 104, except for the last $q$ rows of $P^{-1/2} B^*$, both $B^*$ and $A^*$ perform a convolution on the data. In fact since $\rho_{k+1} \leq \rho_k$, the matrix $KK^*$ is a good rank $N - q$ approximation of $BP^{-1}B^*$. The operations performed on the data vector $\mathbf{y}$ are now clearly seen as a convolution (multiplication by $A^*$), followed by a deconvolution $(A^*)^+$, followed by a convolution $K^*$. If the order of these operations is interchanged, then the matrices $K^*$ and $(A^*)^+$ may be replaced by the appropriately sized matrices $\tilde{K}^*$ and $(\tilde{A}^*)^+$ with $O(p/N)$ error (see Appendix A) and for $p \ll N$, Equation 106 results. This is the ITEMS minimization [31] (modified for damped exponentials) with $k^{th}$ iteration

$$\min_{\mathbf{a}_k} \mathbf{a}_k^* Y \tilde{K} (\tilde{A}_{k-1}^* \tilde{A}_{k-1})^{-1} \tilde{K}^* Y^* \mathbf{a}_k. \tag{108}$$

Equation 107 indicates that the covariance method will also provide a consistent estimate of the exponentials and noise poles. Rewriting the minimization in terms of the polynomials whose coefficients are contained in $\mathbf{a}$ and $\mathbf{b}$ gives

$$\min_{\mathbf{a} \star \mathbf{b}} \| Y (\mathbf{a} \star \mathbf{b}) \|_F^2 \tag{109}$$

where $\star$ indicates the convolution of the polynomial coefficients. The roots of the polynomial whose coefficients are $\mathbf{a} \star \mathbf{b}$ are the exponentials, $\lambda_i$ for $i = 1 \ldots p$, and the noise poles, $\zeta_i$ for $i = 1 \ldots q$. The difficulty is choosing which roots are exponentials and which roots are noise. The amplitude

of exponentials is estimated from Equation 74 with $R_{\mathbf{ww}} = I$ as

$$\mathbf{s} = (H^* H)^{-1} H^* \mathbf{y} = H^+ \mathbf{y}. \tag{110}$$

If $H$ contains a Vandermonde column for each root of the order $p+q$ AR polynomial, then for high signal-to-noise ratios (SNRs) the $p$ roots with the largest amplitude will be the exponentials. Equation 110 provides a good indicator of signal or noise down to near zero SNR.

*5.2.4.6 Choosing Signal and Noise poles.* To see why picking roots with the largest amplitude is a reasonable estimate of the signal, consider the maximum likelihood minimization for white noise

$$\min_{\lambda} \frac{1}{N} \mathbf{y}^* \left( I - G(G^*G)^{-1} G^* \right) \mathbf{y} \tag{111}$$

$$= \max_{\lambda} \frac{1}{N} \mathbf{y}^* G (G^*G)^{-1} G^* \mathbf{y} \tag{112}$$

$$= \max_{\lambda} \frac{1}{N} \mathbf{y}^* G (G^*G)^{-1} G^* G (G^*G)^{-1} G^* \mathbf{y}. \tag{113}$$

Substituting back in the ML estimate of the amplitude gives the optimization

$$\max_{\lambda} \frac{1}{N} \mathbf{s}^* G^* G \mathbf{s}. \tag{114}$$

Thus the ML estimate maximizes the energy of the estimated signal. Now, consider the problem of choosing the signal and noise poles. The function in Equation 114 could be evaluated at all $\binom{p+q}{p}$ possible signals pole combinations and the combination that produced the maximum value chosen. A similar technique could also be employed with the colored noise ML optimization. However, both these methods are computationally intractable for large $p$ or $q$. Thus, a suboptimal approach is needed. Such an approximation to Equation 114 attained by ignoring the cross terms between signals is

$$\max_{\lambda} \frac{1}{N} \sum_{i=1}^{p} s_i^* \mathbf{g}_i^* \mathbf{g}_i s_i, \tag{115}$$

or pick the $p$ poles with the highest energy. For undamped exponentials $\mathbf{g}^* \mathbf{g} = N$, and Equation 115 reduces to

$$\max_{\lambda} \sum_{i=1}^{p} s_i^* s_i, \tag{116}$$

or choose the poles with the largest amplitude. For damped exponentials the appropriate energy measure is

$$\max_{\lambda} \frac{1}{N} \sum_{i=1}^{p} |\lambda_i|^{N-1} s_i^* s_i. \tag{117}$$

Note that while these approaches are suboptimal, they still solve an optimization problem in their own right; the maximum likelihood estimate of a single exponential signal. Thus, at low SNR when the poles of a signal consisting of closely spaces exponentials have merged this technique provides better estimates of the merged pole.

Better performance at low SNR can also be attained with an initial estimate of the $p+q$ exponentials and noise poles from IQML as

$$\min_{\mathbf{a}\star\mathbf{b}} \left\| (\tilde{K}^* A^*)^+ Y(\mathbf{a}\star\mathbf{b}) \right\|_F^2. \tag{118}$$

Equation 109 also suggests that a comparable performance will be attained by increasing the model order of IQML (IQML2) to model both signal poles and noise poles, in effect, estimating an AR($p+q$) process. Results show that this is true, however, picking $q$ poles as noise and whitening the data with Equation 89 still achieves better performance. That is, whitening the noise poles over all frequencies with Equation 89 provides better performance than estimating (eliminating from the signal) the subspace of the noise's peak frequencies by estimating an AR($p+q$) process.

*5.2.4.7 The ARUCIQML Algorithm.* This algorithm will be called Autoregressive Unknown Colored noise IQML (ARUCIQML) and is summarized as follows:

Step 1. *Estimate p exponentials and q noise poles with IQML.* Estimate $\gamma_i$ for $i = 1 \ldots p+q$ and form the $N \times p+q$ matrix $H$.

Step 2. *Determine which $\gamma_i$ are exponentials and which are noise poles.* Estimate $\mathbf{s}$ as

$$\mathbf{s} = H^+ \mathbf{y}. \tag{119}$$

For high signal-to-noise ratios (SNRs) the $p$ exponentials of $H$ with the largest amplitudes are the signal exponentials, $\lambda_i$ for $i = 1 \ldots p$; the remaining $\gamma_i$ are the noise poles, $\zeta_i$ for $i = 1 \ldots q$.

Step 3. *Construct A and L.* Form the coefficients of $a(\lambda)$ and $A$ from the $\lambda_i$ for $i = 1 \ldots p$. Form the coefficients of $b(\zeta) = b_0 + b_1\zeta + \ldots + b_q\zeta^q$ from the $\zeta_i$ for $i = 1 \ldots q$. Form the autocorrelation

sequence of the noise from the AR PSD. Then form $R_{\mathbf{ww}}$ and its Cholesky decomposition $R_{\mathbf{ww}} = L^*L$.

Step 4. *Estimate new* $\lambda_i$ for $i = 1 \ldots p$. Iterate once

$$\min_{\mathbf{a}_k} \left\| (A_{k-1}^* L^*)^+ Y \mathbf{a}_k \right\|_F^2 . \tag{120}$$

Step 5. Form $A$ from the last estimate of $\mathbf{a}$ and repeat Step 4 for several iterations, or until estimates do not significantly change.

Step 6. *Estimate* $\mathbf{s}$ as

$$\hat{\mathbf{s}} = ((L^*)^+ G)^+ (L^*)^+ \mathbf{y}. \tag{121}$$

*5.2.4.8 Autoregressive Moving Average Colored Noise IQML (ARMAIQML).* The following solution for the unknown colored noise problem assumes an ARMA model for the noise (i.e. the noise contains poles and zeros). If the noise sequence is the stationary output of a filter with an equal number of poles and zeros then the noise satisfies an autoregressive moving average (ARMA) difference equation

$$\begin{aligned} b_0 w[n] &= -b_1 w[n-1] \ldots - b_q w[n-q] \\ &\quad + c_0 \epsilon[n] + c_1 \epsilon[n-1] \ldots + c_q \epsilon[n-q], \end{aligned} \tag{122}$$

where $\epsilon[n]$ is a white noise sequence. The ARMA PSD is

$$P_{ww}(e^{j\omega}) = \sigma^2 \frac{\left| \sum_{n=0}^{q} c_n e^{-jn\omega} \right|^2}{\left| \sum_{n=0}^{q} b_n e^{-jn\omega} \right|^2}. \tag{123}$$

Damped exponentials are a deterministic ARMA process. Thus, IQML is again used to simultaneously estimate the exponentials and the noise poles. The exponentials and noise poles are estimated and selected by amplitude as in ARUCIQML. ARUCIQML also estimates the exponential and noise pole amplitudes from Equation 92. However, since the noise amplitudes are stochastic, the numerator coefficients of the ARMA polynomial can not be determined from the noise amplitudes. Instead, the signal and noise are separated as in UCIQML. The noise is filtered (using the noise

poles estimated with IQML) to attain the MA part (numerator) of the noise estimate implied by the IQ̊ML pole estimate.

As in ARUCIQML, IQML provides consistent estimates of the exponentials and noise poles and approximates the maximum likelihood solution (see Appendix B). Similarly to [36] the ARMA covariance matrix is constructed as

$$R_{ARMA} = C^* R_{AR} C \tag{124}$$

where $R_{AR}$ is an $N+p \times N+p$ AR covariance matrix and $C$ is the $N+p \times N$ convolution matrix of the MA (numerator) polynomial,

$$C = \begin{bmatrix} 1 & & 0 \\ c_1 & \ddots & \\ \vdots & \ddots & 1 \\ c_p & & c_1 \\ & \ddots & \vdots \\ 0 & & c_p \end{bmatrix}. \tag{125}$$

As in the AR case, the $N-p \times N-p$ lower triangular Toeplitz matrix $(\tilde{K})^+$ used to approximate $R = (\tilde{K}\tilde{K}^*)^{-1}$ commutes with $C$ and $A$ (see Appendix A and B). The ARMA optimization can thus be stated as $\min \mathbf{y}^* A \tilde{K} W \tilde{K}^* A^* \mathbf{y}$ with a positive definite weighting matrix $W$ and IQML will consistently estimate the exponentials and noise poles.

First for the ARMA noise, the noise and signal are estimated as in ARUCIQML. Then, the signal and noise are separated as in UCIQML as

$$\mathbf{w} = \mathbf{y} - G\mathbf{s} \tag{126}$$

where $G$ contains Vandermonde columns of the exponentials estimated by ARUCIQML and $\mathbf{s}$ is estimated by

$$\mathbf{s} = G^+ \mathbf{y}. \tag{127}$$

The noise poles were also estimated and the AR covariance matrix $R_{AR} = L_{AR}^* L_{AR}$ is constructed from these. The estimated noise sequence $\mathbf{w}$ is then whitened to give the estimated MA noise sequence

$$\mathbf{w}_{MA} = (L_{AR}^*)^+ \mathbf{w}. \tag{128}$$

The autocorrelation sequence of the MA coefficients is then the central $2q+1$ values of the autocorrelation sequence of $\mathbf{w}_{MA}$. The ARMA covariance matrix $R_{\mathbf{ww}}$ is constructed from autocorrelation sequence, the inverse Fourier transform of the ARMA PSD

$$P_{ww}(e^{j\omega}) = \frac{\sum_{m=-q}^{q} r_{MA_m} e^{-j\omega m}}{\left|\sum_{m=0}^{q} b_n e^{-j\omega m}\right|^2}. \tag{129}$$

With the estimate of $\mathbf{a}$ from the signal poles and $L$ from the Cholesky decomposition of $R_{\mathbf{ww}}$, the iteration in Equation 89 is computed. Again, since this iteration improves the estimate of $\mathbf{a}$, and consequently the estimate of $\mathbf{s}$, $\mathbf{s}$ is reestimated at each iteration by Equation 127, and $\mathbf{w}$ is reestimated by Equation 126 giving a new MA autocorrelation sequence $r_{MA}$ and noise covariance matrix $R_{\mathbf{ww}}$.

*5.2.4.9 The ARMAIQML Algorithm.* This algorithm will be called Autoregressive Moving Average colored noise IQML (ARMAIQML) and is summarized as follows:

Step 1. *Estimate $p$ exponentials and $q$ noise poles with IQML.* Estimate $\gamma_i$ for $i = 1 \ldots p+q$ and form the $N \times p+q$ matrix $H$.

Step 2. *Determine which $\gamma_i$ are exponentials and which $\gamma_i$ are noise poles.* Estimate $\mathbf{s}$ as

$$\mathbf{s} = H^+ \mathbf{y}. \tag{130}$$

For high signal-to-noise ratios (SNRs) the $p$ exponentials of $H$ with the largest amplitudes are the signal exponentials, $\lambda_i$ for $i = 1 \ldots p$. The remaining $\gamma_i$ are the noise poles, $\zeta_i$ for $i = 1 \ldots q$.

Step 3. *Construct $G$, $A$, and $L_{AR}$.* Form the coefficients of $a(\lambda)$, and $A$ from the $\lambda_i$ for $i = 1 \ldots p$. Form the coefficients of $b(\zeta) = b_0 + b_1\zeta + \ldots + b_q\zeta^q$ from the $\zeta_i$ for $i = 1 \ldots q$. Form the autocorrelation sequence of the AR part of the noise from the AR PSD. Form the Toeplitz matrix $R_{AR}$ and its Cholesky decomposition $R_{AR} = L_{AR}^* L_{AR}$.

Step 4. *Estimate the MA part of the noise sequence* $\mathbf{w}_{MA}$. Filter out the AR part of the noise by

$$\mathbf{w}_{MA} = (L_{AR}^*)^+(\mathbf{y}-G\mathbf{s}). \tag{131}$$

Step 5. *Estimate the autocorrelation sequence of the MA part of the noise* $r_{MA}[m]$ for $m = -(N-1)\ldots N-1$.

Step 6. *Estimate the autocorrelation sequence of the noise* $r_{\mathbf{ww}}[m]$ for $m = -(N-1)\ldots N-1$. Retain the central $2q+1$ values of the $r_{MA}$ autocorrelation sequence, then calculate the autocorrelation sequence of the noise $r_{\mathbf{ww}}$ from the Fourier transform of the ARMA PSD in Equation 129. Form $R_{\mathbf{ww}}$ and its Cholesky decomposition $R_{\mathbf{ww}} = L^*L$.

Step 7. *Estimate new* $\lambda_i$ for $i = 1\ldots p$. Form $A$ from the last estimate of $\mathbf{a}$ and iterate once

$$\min_{\mathbf{a}_k} \left\| (A_{k-1}^* L^*)^+ Y \mathbf{a}_k \right\|_F^2 . \tag{132}$$

Step 8. *Estimate* $\mathbf{s}$ as

$$\hat{\mathbf{s}} = ((L^*)^+ G)^+ (L^*)^+ \mathbf{y}. \tag{133}$$

Step 9. *Repeat* Step 4 - 8 for several iterations, or until estimates do not significantly change.

*5.2.5 Results.* To test the algorithms in previous section, instances of two exponentials in AR and ARMA colored noise were created. For the first example, the exponential parameters were $(s,\lambda)_i = \{(1, 0.95e^{j0.6\pi})_1, (1, 0.95e^{j0.62\pi})_2\}$. The noise was a two pole process with poles $\zeta = \{0.95e^{j0.2\pi}, 0.95e^{j0.4\pi}\}$. The PSD of the signal and noise are shown in Figure 14. The Mean Square Error (MSE) in estimating the exponentials is shown in Figure 15. The MSE's were calculated for length $N = 32$ snapshots using Monte Carlo simulations each with 200 independent experiments. All of the algorithms were run for ten iterations. The straight solid line in the figures is the Cramér-Rao estimation Bound (CRB) which is described in Appendices C and D.

IQML, IQML with double the known model order (IQML2), UCIQML and CIQML were tested with this data. The results show that CIQML follows the CRB. The MSE of UCIQML will approached the CRB asymptotically as the length of the sample approaches infinity. For the

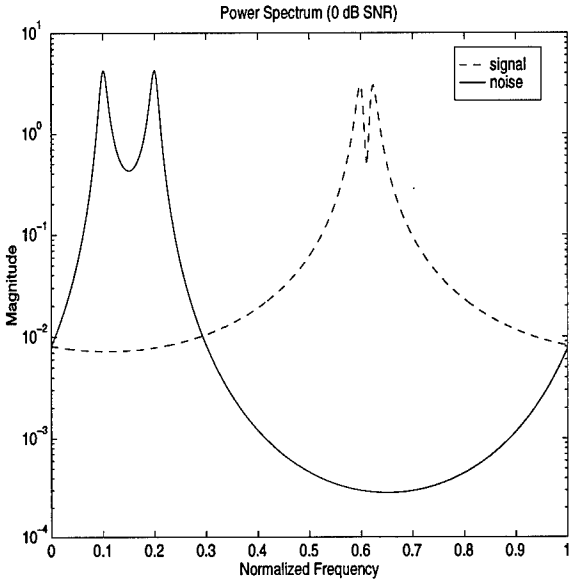


Figure 14. PSD of two exponentials well separated in frequency from AR noise poles (Example 1)

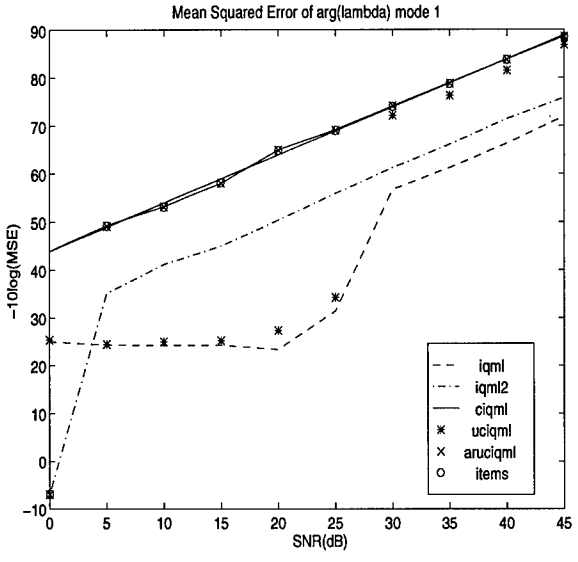


Figure 15. MSE plot for one of two exponentials separated from AR noise (Deterministic ML, Example 1). Techniques: iterative quadratic maximum likelihood (iqml), iqml with double the model order (iqml2), known colored noise iqml (ciqml), unknown colored noise iqml (uciqml), autoregressive uciqml (aruciqml), and iterative estimation of mixed spectra (items)

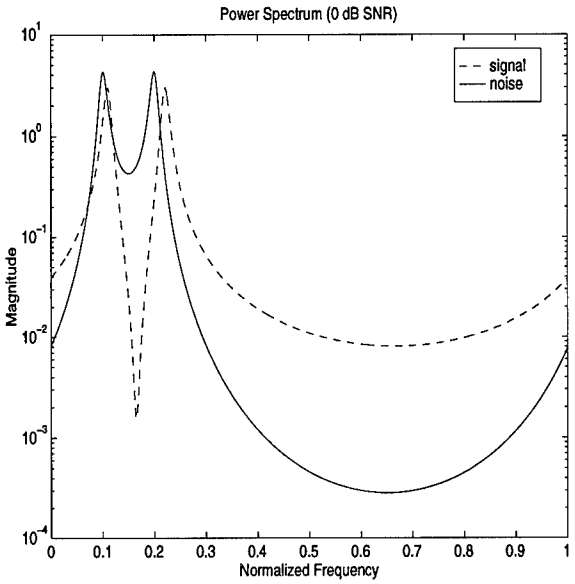


Figure 16. PSD of two exponentials centered near in AR noise poles (Example 2)

length 32 sample, it is quite close to the CRB. IQML attains the CRB for white noise (i.e. the CRB for colored noise is higher than the CRB for white noise). The performance of both IQML and UCIQML (because it is based on the signal estimates of IQML) breaks down at a moderate signal-to-noise ratio (30 dB). The performance of IQML2 does not experience this break-down until much lower SNR. This suggests the techniques exploited in ARUCIQML. The initial estimate for ARUCIQML is attained in the same manner as the IQML2 estimate, which estimates both the signal poles and the noise poles with its increased model order. The ARUCIQML algorithm and the approximate ML method ITEMS [29] [31] described by Equation 106 were also tested on this first example. The exponentials and noise poles were estimated simultaneously with these two algorithms with good results. Their performance matches that of CIQML and tracks the CRB.

For the second AR example, the exponential parameters were changed to $(s, \lambda) = \{(1, 0.95e^{j0.22\pi}), (1, 0.95e^{j0.44\pi})\}$. These exponentials are slightly offset in frequency from the noise poles at $\zeta = (0.95e^{j0.2\pi}, 0.95e^{j0.4\pi})$ and provide a more difficult estimation problem. The PSD of this example is shown in Figure 16. Results are shown in Figure 17. In this example the increased model order of IQML2 was detrimental. IQML2 performance is nonexistent and ITEMS performance is poor. Using ARUCIQML in both AR examples results in improved or comparable performance with IQML.

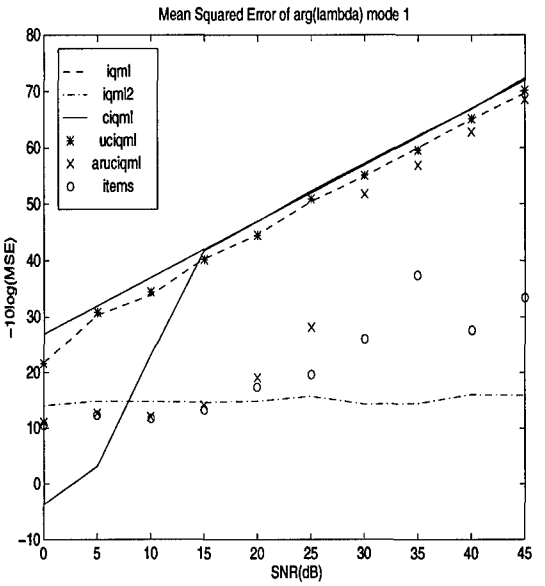


Figure 17. MSE plot for one of two exponentials centered in AR noise (Deterministic ML, Example 2). Techniques: iterative quadratic maximum likelihood (iqml), iqml with double the model order (iqml2), known colored noise iqml (ciqml), unknown colored noise iqml (uciqml), autoregressive uciqml (aruciqml), and iterative estimation of mixed spectra (items)

For the ARMA example, the exponential parameters were $(s,\lambda)_i = \{(1, 0.95e^{j0.6\pi})_1, (1, 0.95e^{j0.62\pi})_2\}$. The noise was a two-pole/two-zero process with poles $\zeta = \{0.95e^{j0.2\pi}, 0.95e^{j0.4\pi}\}$ and zeros $\mu = \{0.95e^{j0.15\pi}, 0.95e^{j0.71\pi}\}$. The PSD of the signal and noise is shown in Figure 18. The Mean Square Error (MSE) in estimating the exponentials is shown in Figure 19.

Again, CIQML follows the CRB, and UCIQML and ARMAIQML will approach the CRB asymptotically. IQML attains the CRB for white noise. The performance of both IQML and UCIQML (because it is based on the signal estimates of IQML) breaks down at a moderate signal-to-noise ratio (30 dB). The performance ($\text{MSE}^{-1}$) of the unknown colored noise algorithms (UCIQML, ARUCIQML, and ARMAIQML) is bounded above by the performance of CIQML and below by IQML applied to the colored noise. As expected when the model used better matches the test case (ARMAIQML vs. ARUCIQML, ARUCIQML vs. UCIQML), the performance is better. The key benefit of UCIQML, ARUCIQML, and ARMAIQML is that the better (colored noise) CRB that is attained. All the algorithms derived in this section also demonstrate good performance in white noise with exponentials $(s,\lambda)_i = \{(1, 0.95e^{j0.6\pi})_1, (1, 0.95e^{j0.62\pi})_2\}$ as shown in Figure 20.

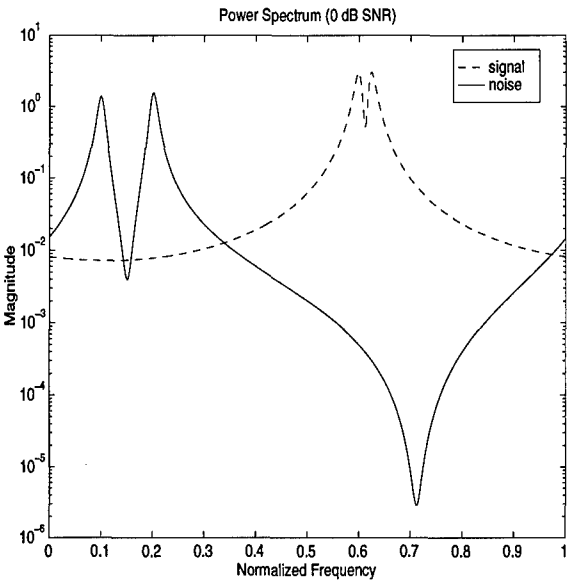


Figure 18. PSD of two exponentials separated from ARMA noise (Example 3)

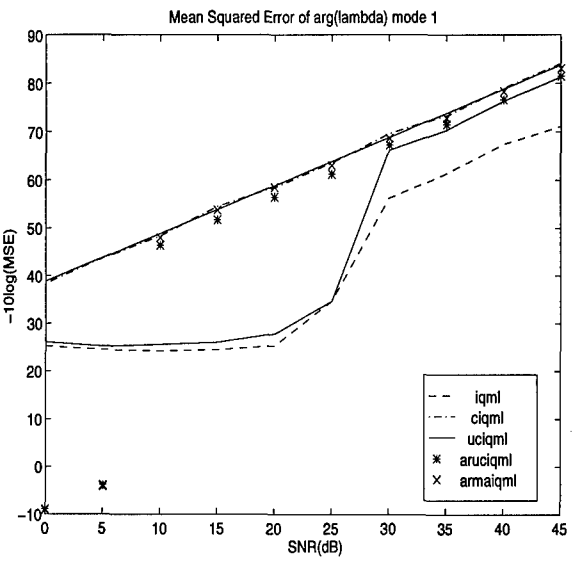


Figure 19. MSE plot for one of two exponentials in ARMA noise (Examples 3). Techniques: iterative quadratic maximum likelihood (iqml), known colored noise iqml (ciqml), unknown colored noise iqml (uciqml), autoregressive uciqml (aruciqml), and autoregressive moving average uciqml (armaiqml)

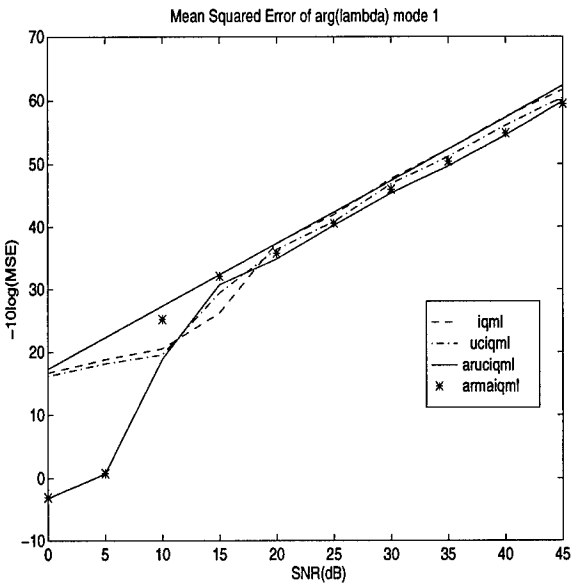


Figure 20. MSE plot one of two exponentials in white noise (Example 4). Techniques: iterative quadratic maximum likelihood (iqml), unknown colored noise iqml (uciqml), autoregressive uciqml (aruciqml), and autoregressive moving average uciqml (armaiqml)

*5.2.6 Conclusions.* Computationally efficient maximum likelihood estimators for colored noise have been developed. These estimators retain all the asymptotic properties of maximum likelihood and produce good results for short data records as well. The ARUCIQML algorithm attains the CRB and is efficient for short data records. The UCIQML, an $N^{th}$ order MA colored noise estimator, and ARMAIQML algorithms do not attain the CRB. This is expected since the colored noise estimates are based on finite order sample covariance estimates which are not efficient for MA and ARMA processes [71].

### *5.3 Stochastic Maximum Likelihood*

This section introduces a new technique to estimate exponentials in colored noise based on stochastic maximum likelihood. This section simplifies the solution for the colored noise stochastic model given in [86] from a global search to an iterative quadratic solution. This simplification uses the projection onto the orthogonal subspace developed in the previous section and the methodology of the white noise stochastic ML method, Method of Direction Estimation (MODE) [69].

The results of [86] allow known and unknown colored noise IQML-like algorithms for the maximum likelihood solution of the stochastic model to be derived. The data in the stochastic

model is distributed as

$$\mathbf{y} \sim N(0, R_{\mathbf{yy}})$$

$$R_{\mathbf{yy}} = GR_{\mathbf{ss}}G^* + \sigma^2 R_{\mathbf{ww}} \tag{134}$$

The general result of [86] finds the maximum likelihood estimate of the unknown parameters with the minimization with respect to the parameters $G$ and $R_{\mathbf{ww}}$ of

$$\mathcal{L}(G, R_{\mathbf{ww}}; \mathbf{y}) = M \ln |P\hat{R}_{\mathbf{yy}}P^* + \hat{\sigma}^2(I-P)R_{\mathbf{ww}}| + MN \tag{135}$$

where $P = L^* P_{L,G} L^{-*}$ and $\hat{\sigma}^2 = (N-p)^{-1}\mathrm{tr}(I-P)\hat{R}_{\mathbf{yy}}R_{\mathbf{ww}}^{-1}$. Observe the term inside the determinate is the estimated covariance matrix separated into signal subspace and noise subspace terms. In terms of the deterministic signal, $\mathbf{y} = G\mathbf{s} + \sigma L^* \epsilon$ and $\hat{R}_{\mathbf{yy}} = \mathbf{yy}^*$ the minimization is

$$\min_{L,G} \ln |(G\mathbf{s} + \sigma G(G^* R_{\mathbf{ww}}^{-1} G)^{-1} G^* L^{-1} \epsilon)(G\mathbf{s} + \sigma G(G^* R_{\mathbf{ww}}^{-1} G)^{-1} G^* L^{-1} \epsilon)^* + \hat{\sigma}^2 (I-P) R_{\mathbf{ww}}|. \tag{136}$$

When the signal and noise are uncorrelated this gives

$$\min_{L,G} \ln |G\mathbf{ss}^* G^* + \sigma^2 G(G^* R_{\mathbf{ww}}^{-1} G)^{-1} G^* L^{-1} \epsilon\epsilon^* L^{-*} G(G^* R_{\mathbf{ww}}^{-1} G)^{-1} G^* + \hat{\sigma}^2 (I-P) R_{\mathbf{ww}}|. \tag{137}$$

Then taking the expected value, $\mathrm{E}\{\cdot\}$, of Equation 137 gives

$$\min_{L,G} \ln |(G\mathbf{ss}^* G^* + \sigma^2 P R_{\mathbf{ww}} + \hat{\sigma}^2 (I-P) R_{\mathbf{ww}}|. \tag{138}$$

The first term in this minimization, the signal, is fixed, thus, the minimization is satisfied when $\sigma^2$ is minimized. Observe that $\sigma^2$ is still minimized when the orthogonal projection $P_{L,A}$ is used in place of $P_{G,A}$. Replacing $P_{G,A}$ with $P_{L,A}$ in Equation 135 and some simplification gives

$$\begin{aligned}
& \min_{L,A} \ln |L^* P_{L,A} L^{-*} \hat{R}_{\mathbf{yy}} (L^* P_{L,A} L^{-*})^* + \hat{\sigma}^2 (I - L^* P_{L,A} L^{-*}) R_{\mathbf{ww}}| \\
= & \min_{L,A} \ln |R_{\mathbf{ww}} A (A^* R_{\mathbf{ww}} A)^{-1} A^* \hat{R}_{\mathbf{yy}} A (A^* R_{\mathbf{ww}} A)^{-1} A^* R_{\mathbf{ww}} \\
& + \hat{\sigma}^2 (I - R_{\mathbf{ww}} A (A^* R_{\mathbf{ww}} A)^{-1} A^*) R_{\mathbf{ww}}| \\
= & \min_{L,A} \ln |R_{\mathbf{ww}} A (A^* R_{\mathbf{ww}} A)^{-1} A^* \hat{R}_{\mathbf{yy}} A (A^* R_{\mathbf{ww}} A)^{-1} A^* R_{\mathbf{ww}} \\
& - \hat{\sigma}^2 R_{\mathbf{ww}} A (A^* R_{\mathbf{ww}} A)^{-1} A^* R_{\mathbf{ww}} A (A^* R_{\mathbf{ww}} A)^{-1} A^* R_{\mathbf{ww}} + \hat{\sigma}^2 R_{\mathbf{ww}}|
\end{aligned}$$

$$
\begin{aligned}
&= \min_{L,A} \ln |R_{\mathbf{ww}} A (A^* R_{\mathbf{ww}} A)^{-1} A^* (\hat{R}_{\mathbf{yy}} - \hat{\sigma}^2 R_{\mathbf{ww}}) A (A^* R_{\mathbf{ww}} A)^{-1} A^* R_{\mathbf{ww}} + \hat{\sigma}^2 R_{\mathbf{ww}}| \\
&= \min_{L,A} \ln |R_{\mathbf{ww}} A (A^* R_{\mathbf{ww}} A)^{-1} A^* (\hat{R}_{\mathbf{yy}} - \hat{\sigma}^2 R_{\mathbf{ww}}) A (A^* R_{\mathbf{ww}} A)^{-1} A^* + \hat{\sigma}^2 I| + \ln |R_{\mathbf{ww}}| \\
&= \min_{L,A} \ln |A (A^* R_{\mathbf{ww}} A)^{-1} A^* (\hat{R}_{\mathbf{yy}} - \hat{\sigma}^2 R_{\mathbf{ww}}) + \hat{\sigma}^2 I| + \ln |R_{\mathbf{ww}}| \quad (|AB + I| = |BA + I|) \\
&= \min_{L,A} \ln |(A^* L^*)^+ A^* (\hat{R}_{\mathbf{yy}} - \hat{\sigma}^2 R_{\mathbf{ww}}) A (LA)^+ + \hat{\sigma}^2 I| + \ln |R_{\mathbf{ww}}| \qquad (139)
\end{aligned}
$$

Notice again that the second term in Equation 139 does not depend on $A$. Thus this term can be eliminated from the minimization to find $A$. Since the logarithm is a monotonic function the minimization in Equation 139 is equivalent to one with the logarithm removed. Then, for a consistent estimate of $R_{\mathbf{ww}}$, $\hat{R}_{\mathbf{ww}}$, the minimization is

$$
\min_{L,A} |(A^* L^*)^+ A^* (\hat{R}_{\mathbf{yy}} - \hat{\sigma}^2 \hat{R}_{\mathbf{ww}}) A (LA)^+ + \hat{\sigma}^2 I| \qquad (140)
$$

Note the quantity in the determinant is positive semidefinite. Then, the inequality

$$
|R| \leq (\frac{\mathrm{tr} R}{N})^N \qquad (141)
$$

for $R$ positive semidefinite [41] indicates that the minimization in Equation 140 is equivalent to

$$
\min_{L,A} \mathrm{tr} \left( (A^* L^*)^+ A^* (\hat{R}_{\mathbf{yy}} - \hat{\sigma}^2 \hat{R}_{\mathbf{ww}}) A (LA)^+ + \hat{\sigma}^2 I \right) \qquad (142)
$$

For white noise this is equivalent to the weighted subspace minimization given in [50]. It is shown in [71] that the stochastic ML model may be applied with consistency to deterministic data. In the form of Equation 142, the effect of the minimization on the signal, $\mathbf{y} = G\mathbf{s} + \sigma L^* \epsilon$, is to minimize

$$
\begin{aligned}
&\mathrm{E}\{\mathrm{tr} \left( (A^* L^*)^+ A^* (\hat{R}_{\mathbf{yy}} - \hat{\sigma}^2 \hat{R}_{\mathbf{ww}}) A (LA)^+ + \hat{\sigma}^2 I \right)\} \\
&= \mathrm{E}\{\mathrm{tr} \left( (A^* L^*)^+ A^* G \mathbf{s}\mathbf{s}^* G^* A (LA)^+ + \sigma^2 (A^* L^*)^+ A^* L^* \epsilon\epsilon^* L A (LA)^+ - \hat{\sigma}^2 I + \hat{\sigma}^2 I \right)\} \\
&= \mathrm{tr} \left( \sigma^2 (A^* L^*)^+ A^* L^* \mathrm{E}\{\epsilon\epsilon^*\} L A (LA)^+ \right) \\
&= \sigma^2 \mathrm{tr}(P_{L,A}) \\
&= (N - p)\sigma^2 \qquad (143)
\end{aligned}
$$

where $\hat{R}_{\mathbf{ww}} = \mathbf{yy}^*$, $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, and $A^*G = 0$. Thus, the variance of the noise is minimized. Observe that the difference from the data model is the way in which the effect of the noise on the exponential signal estimation (choice of minimizing $A$) is minimized. In the deterministic model, a whitening filter minimizes the noise. In the stochastic model, an estimate of the noise is subtracted from the signal prior to estimation of the minimizing $A$. This is readily accomplished on the whitened data in an extension of the white noise technique MODE [69] to known colored noise which will be called Colored noise MODE (CMODE). MODE is described in Appendix E. Including our previously canceled $L^*L^{-*}$ and $L^{-1}L$, and taking the principal components (pc) of the whitened data (the eigenvectors associated with the $p$ largest eigenvalues) gives the minimization

$$\min_A \mathrm{tr}\left((A^*L^*)^+ A^*L^* \mathrm{pc}\left\{L^{-*}(\hat{R}_{\mathbf{yy}} - \hat{\sigma}^2\hat{R}_{\mathbf{ww}})L^{-1}\right\} LA(LA)^+\right) \tag{144}$$

or

$$\min_A \mathrm{tr}\left(A(A^*\hat{R}_{\mathbf{ww}}A)^{-1}A^*L^* \mathrm{pc}\left\{L^{-*}(\hat{R}_{\mathbf{yy}} - \hat{\sigma}^2\hat{R}_{\mathbf{ww}})L^{-1}\right\} L\right) \tag{145}$$

where $\hat{\sigma}^2$ may be quickly estimated as the average of the remaining $N-p$ eigenvalues. Notice, that since $\hat{\sigma}^2$ is now considered any consistent estimate of $\sigma^2$, and because of the additive property of the trace, the final $\hat{\sigma}^2 I$ term in Equation 142 may be dropped. This, in turn, implies that any weighting of $\hat{\sigma}^2 I$ may be subtracted in Equation 145 so long as the matrix $R$ in Equation 141 remains positive semidefinite. A discussion of the optimum weighting is included in the next section.

Equation 145 is minimized in a similar fashion to IQML by creating a data matrix $Y$ in Equation 87 to form the minimization of a quadratic in $\mathbf{a}$. Here, also, is the first compromise necessary for a single data instance. The full covariance matrix $R_{\mathbf{yy}}$ is not available; $\mathbf{yy}^*$ is rank 1 and a principal components analysis is meaningless. A smaller $(N-p \times N-p)$ rank-$p+1$ covariance matrix $\hat{R}_{\mathbf{yy}} = YY^*$ may be estimated from overlapping segments of the data with $Y$ from Equation 87. In 1-D this only increases the number of samples required. In 2-D estimation this will force a compromise between true 2-D exponential estimation and 1-D by 1-D estimation.

*5.3.1 Estimation in Unknown, AR and ARMA Colored noise.* Stochastic algorithms corresponding to each of the deterministic algorithms were developed and tested on the same set of examples as in Section 4.4. The stochastic algorithms are built around the iteration in Equation 145. An eigendecomposition of the whitened outer product of the data matrix in Equation 87,

$(L^*)^+RL^+ = (L^*)^+YY^*L^+ = VDV^*$, gives $p$ signal eigenvectors $L^*V_pW^{1/2}$ where $V_p$ contains the $p$ columns of $V$ associated with the $p$ largest eigenvalues in $D$, and $W$ is the optimal weighting of the eigenvectors [71] (see Appendix E). The $p$ signal eigenvectors are used in place of the data matrix of the colored noise version of IQML. The best performing and most computationally efficient method of finding principal components involved taking the SVD of $(L^*)^+Y$ . This quickly identifies the eigenvectors $V_p$ and the square roots of the eigenvalues in $D$. Other aspects of the algorithms (how the noise covariance is consistently estimated) are the same as the deterministic versions.

The optimum weight matrix $W$ for a large number of data instances is [50]

$$W = (D - \hat{\sigma}^2 I)^2 D^{-1} = (D - \hat{\sigma}^2 I)(I - \hat{\sigma}^2 D^{-1}). \tag{146}$$

The effect of this weighting matrix can be understood in terms of the inequality in Equation 141 which is derived from the dominance of the arithmetic mean of the positive and zero eigenvalues of the matrix $R$ in Equation 141 over the geometric mean of these eigenvalues. This inequality becomes an equality when the eigenvalues are equal. Then, as $\hat{\sigma}^2 I \rightarrow D$ (low SNR) the weighting term in Equation 146 goes to zero and the approximation of using the trace in Equation 142 becomes more accurate.

The same two mode test cases as in Section 4.4 were run with MODE, and its known colored noise (CMODE) and unknown colored noise variants (UCMODE, ARUCMODE, ARMAMODE). The results are shown in Figures 21 through 24. When multiple instances of the data are available and when the exponentials in the signals are correlated these methods can show improved performance over the IQML based methods [71]. For the single data instance, however, these algorithms do not in general outperform the IQML based methods. In some instances, the MODE based algorithms do show better performance (UCMODE). As in most instances of the deterministic algorithms, the unknown noise versions of the algorithms very closely approach the known colored noise version's performance. The MODE based algorithms, however, all fall slightly below the CRB. This is due the fact that an extra $p$ samples are used to effectively estimate the signal subspace, thus the optimization in Equation 145 is based on a set of length $N - p$ data samples while the deterministic optimization is performed on length $N$ data sample.

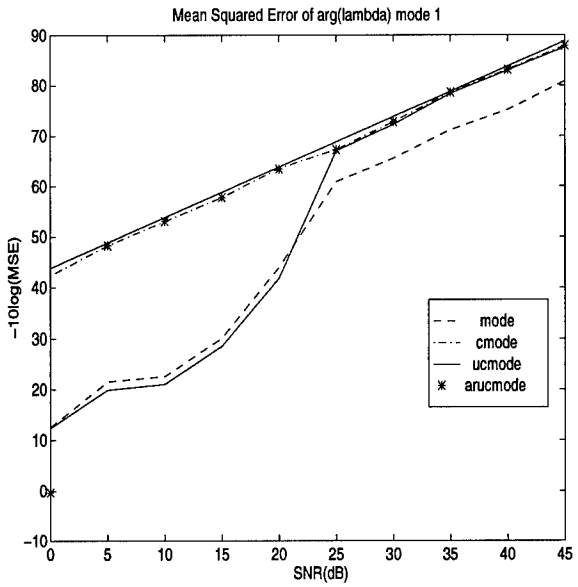


Figure 21. MSE plot for two exponentials widely separated from AR noise (Stochastic ML, Example 1). Techniques: method of direction estimation (mode), known colored noise mode (cmode), unknown colored noise mode (ucmode), and autoregressive ucmode (arucmode)

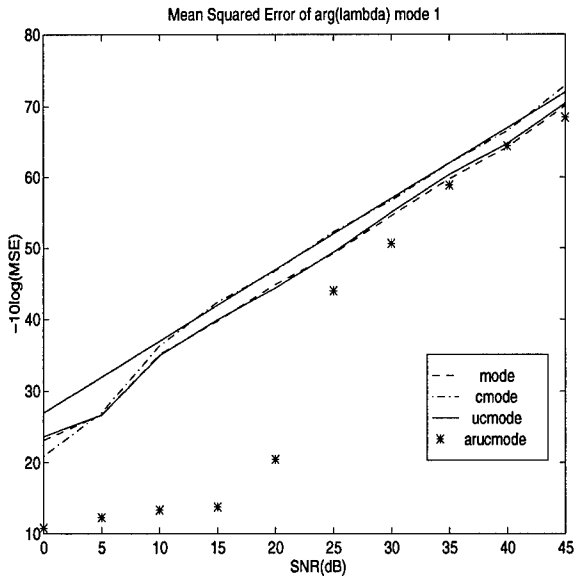


Figure 22. MSE plot for two exponentials centered in AR noise (Stochastic ML, Example 2). Techniques: method of direction estimation (mode), known colored noise mode (cmode), unknown colored noise mode (ucmode), and autoregressive ucmode (arucmode)

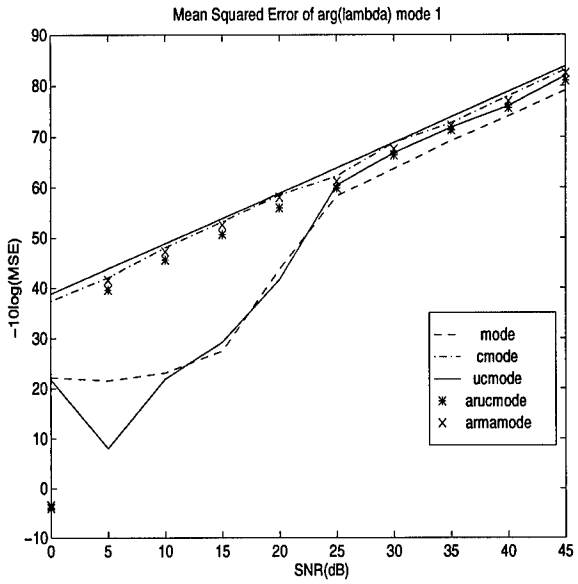


Figure 23. MSE plot for two exponentials in ARMA noise (Stochastic ML, Example 3). Techniques: method of direction estimation (mode), known colored noise mode (cmode), unknown colored noise mode (ucmode), autoregressive ucmode (arucmode), and autoregressive moving average ucmode (armamode)

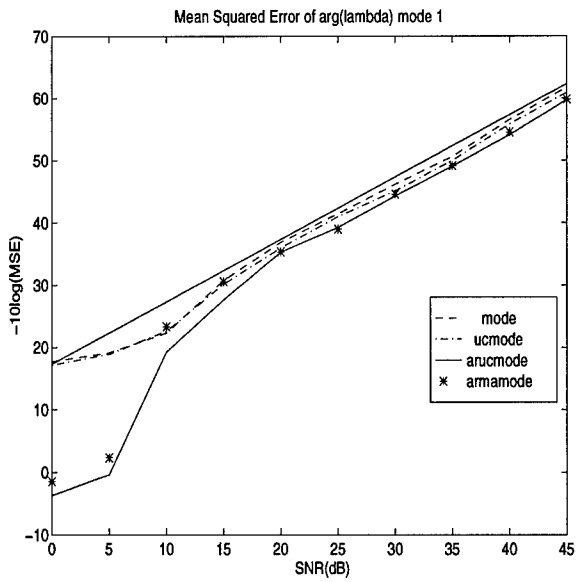


Figure 24. MSE plot for two exponentials in white noise (Stochastic ML, Example 4). Techniques: method of direction estimation (mode), unknown colored noise mode (ucmode), autoregressive ucmode (arucmode), and autoregressive moving average ucmode (armamode)

# *VI. Estimation of 2-D Exponentials in Colored Noise*

## *6.1 Overview*

This chapter develops extensions to two dimensions of the one-dimensional maximum likelihood exponential estimation techniques developed in the previous chapter. This extension is not completely straight forward because of the unique nature of two-dimensional (2-D) signals. Two-dimensional discrete time signal processing theory has been well developed [14], and many of the characteristics follow directly from one-dimensional discrete time signal processing. However, one dimensional spectral estimation techniques that involve polynomial rooting fail to be directly applicable due to the absence of a fundamental theorem of algebra in two dimensions. The roots of a polynomial in two dimensions are not unique and efficient root finding techniques do not exist. This fact has lead to the development of several novel methods of 2-D spectral estimation. This section briefly reviews the 2-D techniques that have been developed in order to understand how the two-dimensional maximum likelihood solution effectively exploits both dimensions. First, to discuss the 2-D techniques, the following signal model of a single instance of data with exponentials in two dimensions is used. Two-dimensional damped exponential data, $y$, in noise, $n$, containing $p$ exponentials may be described as

$$y[m_1, m_2] = \sum_{i=1}^{p} s_i \lambda_i^{m_1} \gamma_i^{m_2} + n[m_1, m_2] \quad 0 \le m_1 \le M_1 - 1 \quad 0 \le m_2 \le M_2 - 1 \tag{147}$$

where the $\lambda_i$ and $\gamma_i$ are the complex exponential frequencies and the $s_i$ are the complex amplitudes of the exponentials. Arranging the data $y$ in an $M_1 \times M_2$ matrix $Y$ gives

$$Y = GSH^T + N, \tag{148}$$

where

$$G = [\ \mathbf{g}_1 \quad \mathbf{g}_2 \quad \cdots \quad \mathbf{g}_p\ ] \quad \mathbf{g}_i = \Psi_{M_1}(\lambda_i), \tag{149}$$

$$S = \mathrm{diag}([\ s_1 \quad s_2 \quad \cdots \quad s_p\ ]) \tag{150}$$

$$H = [\ \mathbf{h}_1 \quad \mathbf{h}_2 \quad \cdots \quad \mathbf{h}_p\ ] \quad \mathbf{h}_i = \Psi_{M_2}(\gamma_i), \tag{151}$$

and

$$\Psi_M(z) = [\ 1 \quad z \quad z^2 \quad \cdots \quad z^M\ ]^T, \tag{152}$$

$N$ is a random field of noise, and $S$ is diagonal. Another less restrictive model of the data is also possible if all the elements of the matrix $S$ are populated. In this case, the model is separable and 1-D techniques can be efficiently applied in each dimension. In this case, exponentials with any combination of spatial frequencies in each dimension are possible (a 1-D by 1-D model). This model then requires a strict pairing of spatial frequencies if two-dimensional exponentials at unique frequencies are present in the data. Conversely, the restricted (full 2-D) model where $S$ is diagonal does not require pairing since its solution is unique.

### *6.2 Two-Dimensional Techniques*

Several two-dimensional exponential estimation techniques are introduced here that follow a 1-D by 1-D, or a 2-D model. A detailed analysis of the performance of these techniques is contained in [52]. The techniques discussed here are those that do not involve search of the spatial frequency plane, but instead find the spectral peaks through some more efficient means.

The first technique from [2] is developed from a state space representation of the data and uses a singular value decomposition of the matrix

$$Y = U\Sigma V^*$$

where $U$ and $V$ are unitary matrixes spanning the column space and row space of $Y$ respectively, and $\Sigma$ is a diagonal matrix of the singular values of $Y$. The matrices $G$, $S$, and $H$ are estimated from $U$, $\Sigma$,and $V$, respectively. The spatial frequencies are estimated separately in each dimension from $U$ and $V$ by exploiting the rotational invariance of exponentials [57]. As in all the remaining techniques requiring pairing, the frequencies are efficiently paired [52] by row and column of the $p$ largest amplitudes in the least squares estimate of the matrix

$$S = G^{+}Y(H^{T})^{+}. \tag{153}$$

This method attains the multi-trial 1-D CRB given in [10].

The second technique, Matrix Enhanced Matrix Pencil (MEMP) from [25] also exploits rotational invariance to find the spatial frequencies. This method also requires pairing, however, it efficiently estimates the 2-D exponentials of the restricted 2-D model by simultaneously using the

data from all the columns or rows, and attains the 2-D CRB [10]. In MEMP, the frequencies are estimated from a block Hankel matrix consisting of blocks which are the Hankel data matrix of the 1-D covariance method [30] for each column of data . This block matrix is in effect an estimate of the Cholesky decomposition of the full $M_1M_2 \times M_1M_2$ covariance matrix of the data, which has been ordered columnwise into one vector. However, since only one data instance is available and some samples are used up to make the matrix rank $p$, the block matrix is not the full size 2-D covariance matrix.

The 2D Prony method [59] is an extension of the 1-D Prony technique [30]. The covariance method is applied simultaneously to all the columns (rows) of data, in effect, considering each column (row) as a new data instance. This method uses the one-dimensional technique of over-modeling (estimating more frequencies than the data contains) and selects the two-dimensional frequency combinations that produce the highest energy signal.

The 2D MODE technique [38] estimates the full 2-D covariance matrix from the outer product of a column-ordered vector of the data (a rank one estimate for the single data instance). Separately averaging of the elements of the covariance matrix associated with one column (row) of $S$ allows full rank covariance matrices for each dimension to be used in the 1-D stochastic maximum likelihood technique, MODE [69]. In a later section, the 2D MODE technique is extended to create colored noise techniques based on stochastic maximum likelihood.

For high signal-to-noise ratios (SNRs), all of these techniques perform well with respect to the appropriate CRB. Of these diverse techniques, 2D MODE [38] and 2D IQML [10] (discussed in detail later) can be related to a maximum likelihood solution of the two-dimensional exponential estimation problem in colored noise, and only MEMP [25] and 2D IQML [10] achieve the 2-D CRB [52].

### *6.3 Deterministic Maximum Likelihood*

In this section, the two-dimensional maximum likelihood problem is formulated, then, via this formulation the techniques are extended to the unknown colored noise problem. The 2-D maximum likelihood solution follows almost directly from 1-D solution when the columns of the data matrix

are stacked (this is the *vec*torize operation). Equation 148 then becomes

$$\mathbf{y} = \text{vec}(Y) = (H \otimes G)\mathbf{s} + \mathbf{n} \tag{154}$$

or

$$\mathbf{y} = F\mathbf{s} + \mathbf{n} \quad \mathbf{y} \sim \text{N}(F\mathbf{s}, \sigma^2 R_{\mathbf{nn}}) \tag{155}$$

where $\otimes$ denotes the Kronecker product, $\mathbf{s} = \text{vec}(S)$, and $\mathbf{n} = \text{vec}(N)$. Equation 148 implies that the 2-D harmonics lie on a 1-D by 1-D grid. Removing this restriction by forcing $S$ to be diagonal eliminates the elements of $\mathbf{s}$ in equation 154 where $s_i = 0$. The new $M_1 M_2 \times p$ matrix $F$ becomes

$$F = [\ \mathbf{h}_1 \otimes \mathbf{g}_1 \quad \mathbf{h}_2 \otimes \mathbf{g}_2 \quad \cdots \quad \mathbf{h}_p \otimes \mathbf{g}_p\ ] \tag{156}$$

where $\mathbf{s}$ now contains just the diagonal elements of $S$. Additionally, the noise is circular Gaussian such that $\text{E}\{\mathbf{n}^*\mathbf{n}\} = \sigma^2 R_{\mathbf{nn}}$ and $\text{E}\{\mathbf{n}^T\mathbf{n}\} = 0$, where $(\cdot)^*$ denotes complex conjugate transpose.

Analogous to the 1-D case, the probability density and log-likelihood functions are

$$f(\mathbf{y}) = \frac{1}{|\pi\sigma^2 R_{\mathbf{nn}}|} e^{-\frac{1}{\sigma^2}(\mathbf{y}-F\mathbf{s})^* R_{\mathbf{nn}}^{-1}(\mathbf{y}-F\mathbf{s})}, \tag{157}$$

and

$$\mathcal{L}(F, \mathbf{s}, \sigma^2 R_{\mathbf{nn}}; \mathbf{y}) = -M_1 M_2 \ln \pi\sigma^2 - \ln |R_{\mathbf{nn}}| - \frac{1}{\sigma^2}(\mathbf{y} - F\mathbf{s})^* R_{\mathbf{nn}}^{-1}(\mathbf{y} - F\mathbf{s}). \tag{158}$$

The zeros of $\frac{\delta\mathcal{L}}{\delta\sigma^2}$, and $\frac{\delta\mathcal{L}}{\delta\mathbf{s}^*}$ are the maximum likelihood (ML) estimates of $\sigma^2$ and $\mathbf{s}$,

$$\hat{\sigma}^2 = \frac{1}{M_1 M_2}(\mathbf{y} - F\mathbf{s})^* R_{\mathbf{nn}}^{-1}(\mathbf{y} - F\mathbf{s}) \tag{159}$$

$$\hat{\mathbf{s}} = (F^* R_{\mathbf{nn}}^{-1} F)^{-1} F^* R_{\mathbf{nn}}^{-1}\mathbf{y} = L^* P_{L,F} L^{-*}\mathbf{y} \tag{160}$$

where the positive definite covariance matrix is decomposed as $R_{\mathbf{nn}} = L^* L$ and $P_{L,F} = L^{-*} F(F^* R_{\mathbf{nn}}^{-1} F)^{-1} F^* L^{-1}$ where $L^{-*} = (L^*)^{-1}$. Substituting Equations 159 and 160 into Equation 158 gives the compressed log likelihood function

$$\mathcal{L}(F, R_{\mathbf{nn}}; \mathbf{y}) = -M_1 M_2 \ln \frac{\pi}{M_1 M_2}\mathbf{y}^*(I - L^* P_{L,F} L^{-*})^* R_{\mathbf{nn}}^{-1}(I - L^* P_{L,F} L^{-*})\mathbf{y} - \ln |R_{\mathbf{nn}}| - M_1 M_2. \tag{161}$$

The maximum likelihood estimates of the remaining parameters are then found by maximizing $\mathcal{L}(G, R_{\mathbf{nn}}; \mathbf{y})$ or equivalently minimizing

$$\begin{aligned} & \min_{L,F} \ln |\mathbf{y}^* L^{-1}(I - P_{L,F}) L^{-*} \mathbf{y} R_{\mathbf{nn}}| \\ = & \min_{L,W} \ln |\mathbf{y}^* L^{-1} P_{L,W} L^{-*} \mathbf{y} R_{\mathbf{nn}}| \end{aligned} \tag{162}$$

where the orthogonal projection to $P_{L,F}$ is

$$P_{L,W} = LW(W^* R_{\mathbf{nn}} W)^{-1} W^* L^* \tag{163}$$

where the column space of $W$ spans the subspace orthogonal to the signal subspace, $W^* F = 0$ and $W$, or its eigenspace equivalent is full column rank. Then using the Cholesky decomposition of $R_{\mathbf{nn}} = L^* L$, and full column rank and full row rank definitions of the Moore-Penrose pseudoinverse, $A^+ = (A^* A)^{-1} A^*$ and $A^+ = A^*(AA^*)^{-1}$ respectively,

$$\begin{aligned} P_{L,W} &= LW(W^* R_{\mathbf{nn}} W)^{-1} W^* R_{\mathbf{nn}} W (W^* R_{\mathbf{nn}} W)^{-1} W^* L^* && (164) \\ &= LW(W^* R_{\mathbf{nn}} W)^{-1} W^* L^* LW (W^* R_{\mathbf{nn}} W)^{-1} W^* L^* && (165) \\ &= LW(LW)^+ (W^* L^*)^+ W^* L^* && (166) \\ &= LW(W^* R_{\mathbf{nn}} W)^+ W^* L^*. && (167) \end{aligned}$$

Thus, the projection can be constructed from $W$, even if $W$ is not full row or column rank, as is the case in the following section.

Again as in the 1-D case the minimization in Equation 162 may be written as

$$\min_{L,W} \ln |\mathbf{y}^* L^{-1} P_{L,W} L^{-*} \mathbf{y}| + \ln |R_{\mathbf{nn}}| \tag{168}$$

where parameters in $W$ can be maximized independently of the $\ln |R_{\mathbf{nn}}|$ term. Thus, given a consistent estimate of $R_{\mathbf{nn}}$ , $\hat{R}_{\mathbf{nn}}$ the parameters in $W$ that minimize Equation 168 are found from

$$\min_{W} \mathbf{y}^* W (W^* \hat{R}_{\mathbf{nn}} W)^+ W^* \mathbf{y}. \tag{169}$$

*6.3.1 Known Colored Noise Maximum Likelihood.* A matrix $W$ that spans the subspace orthogonal to the signal subspace is derived in [9] as the $((2M_1-1)(M_2-p)+M_1-1)\times M_1M_2$ matrix

$$W^* = \begin{bmatrix} I_{M_1} \otimes C_p^* \\ [I_{M_1}|0] \otimes D_{p-1}^* - [0|I_{M_1}] \otimes \hat{I} \end{bmatrix}, \tag{170}$$

where $\hat{I} = [\ I_{M_1-p+1} \quad \mathbf{0}_1 \quad \mathbf{0}_2 \quad \cdots \quad \mathbf{0}_{p-1}\ ]$ and the $(M_2-p)\times M_2$ matrix $C_p^*$, and $(M_2-p+1)\times M_2$ matrix $D_{p-1}^*$ are defined from

$$C_k^* = \begin{bmatrix} c_k & \cdots & c_1 & c_0 & 0 \\ & \ddots & & \ddots & \ddots \\ 0 & c_k & \cdots & c_1 & c_0 \end{bmatrix}, \tag{171}$$

and

$$D_{k-1}^* = \begin{bmatrix} d_{k-1} & \cdots & d_1 & d_0 & 0 \\ & \ddots & & \ddots & \ddots \\ 0 & d_{k-1} & \cdots & d_1 & d_0 \end{bmatrix}. \tag{172}$$

The upper block of $W^*$ annihilates the harmonics in one direction $\lambda_i$, and the lower block of $W^*$ predicts the harmonics in the second dimension $\gamma_j$, based on the harmonics annihilated in the first dimension. The minimization is performed with IQML-like iterations,

$$\min_{\mathbf{f}_i} \mathbf{f}_i^* \tilde{Y}^* (W_{i-1}^* \hat{R}_{\mathbf{nn}} W_{i-1})^+ \tilde{Y} \mathbf{f}_i, \tag{173}$$

where $\mathbf{f} = [\ \mathbf{c}^T \quad \mathbf{d}^T\ ]^T$, $\mathbf{c} = [\ c_0 \quad c_1 \quad \cdots \quad c_k\ ]^T$, $\mathbf{d} = [\ d_0 \quad d_1 \quad \cdots \quad d_{k-1}\ ]^T$, and $\tilde{Y}$ is defined such that

$$\tilde{Y}\mathbf{f} = W^*\mathbf{y}. \tag{174}$$

The pseudoinverse in Equation 173 requires a significant amount of memory. An new equivalent implementation based on the relation

$$(W^*W)^+ = ((WW^*)^+W)^*(WW^*)^+W \tag{175}$$

developed in Appendix F, requires only one quarter the memory. Then, the minimization in Equation 173 becomes

$$\min_{\mathbf{f}_i} \left\| ((LW_{i-1}W_{i-1}^* L^*)^+ LW_{i-1})\tilde{Y}\mathbf{f}_i \right\|_F^2 \tag{176}$$

Computations can be reduced further from $O(M^6)$ ($M = M_1 = M_2$; $p \ll M$) for the SVD in $(W^*W)^+$ when this is replaced with the normal inverse by using a permutation matrix. The matrix $P$ chooses $M_1M_2 - p$ linearly independent rows of the banded matrix $W^*$ [12]. With the $(M_1M_2-p) \times (M_1M_2-p)$ matrix $PW^*WP^*$ Equation 173 becomes

$$\min_{\mathbf{f}_{i+1}} \mathbf{f}_{i+1}^* \tilde{Y}^* P^* (PW_i^* R_{\mathbf{nn}} W_i P^*)^{-1} P\tilde{Y}\mathbf{f}_{i+1}. \tag{177}$$

Each of the Kronecker products in the definition of $W^*$ is a upper band matrix with the last having the largest bandwidth $M$. Thus there exists a matrix $PW^*$ with bandwidth $M$ and a band Cholesky solution to Equation 177 is $O(M^4)$ [19].

*6.3.2 2-D Unknown Colored Noise Maximum Likelihood.* This section develops a new technique to estimate 2-D exponentials in unknown colored noise by extending the 2D IQML technique. When the noise is white, $R_{\mathbf{nn}} = I$, the algorithm to perform the minimization in Equation 173 is called 2D IQML. This algorithm achieves the 2D CRB [9] [11] [52]. As in the 1-D case, due to the Wold decomposition the 2-D CRB for the exponential parameters is the same for unknown colored noise and known colored noise. Also as in the 1-D case, $R_{\mathbf{nn}}$ can be replaced with a consistent estimate, $\hat{R}_{\mathbf{nn}}$, and the asymptotic properties of maximum likelihood minimization in Equation 168 will be retained. The unknown colored noise variants for 2-D now follow directly form the 1-D cases.

If the noise sequence is assumed stationary then the noise sequence is

$$\mathbf{n} = \mathbf{y} - F\mathbf{s} \tag{178}$$

where $F$ is estimated by 2D IQML and $\mathbf{s}$ is estimated by

$$\mathbf{s} = (F^*F)^{-1}F^*\mathbf{y} = F^+\mathbf{y}. \tag{179}$$

Then for stationary noise, $R_{\mathbf{nn}}$ is the block Toeplitz covariance matrix [28] whose elements are the autocorrelation coefficients of $N$ in Equation 148. $N$ is attained by reversing the vec operation and extracting length $M_1$ segments of the vector $\mathbf{n}$ and ordering them into a matrix. The autocorrelation coefficients of $N$ are then the $2M_1 - 1 \times 2M_2 - 1$ matrix $\hat{r}_{NN}$. The matrix $\hat{r}_{NN}$ is quickly attained by taking the magnitude squared of the two-dimensional FFT of $N$ zero-padded to $2M_1 - 1$ by $2M_2 - 1$. Then the 2-D autocorrelation matrix $R_{\mathbf{nn}}$ is the block Toeplitz matrix

$$R_{\mathbf{nn}} = \begin{bmatrix} R_0 & R_{-1} & \cdots & R_{-M_2} \\ R_1 & R_0 & & R_{-(M_2-1)} \\ \vdots & & \ddots & \\ R_{M_2} & R_{M_2-1} & & R_0 \end{bmatrix} \tag{180}$$

where

$$R_i = \begin{bmatrix} \hat{r}_{NN}[0,i] & \hat{r}_{NN}[-1,i] & \cdots & \hat{r}_{NN}[-M_1,i] \\ \hat{r}_{NN}[1,i] & \hat{r}_{NN}[0,i] & & \hat{r}_{NN}[-(M_1-1),i] \\ \vdots & & \ddots & \\ \hat{r}_{NN}[M_1,i] & \hat{r}_{NN}[M_1-1,i] & & \hat{r}_{NN}[0,i] \end{bmatrix}. \tag{181}$$

With the estimate of $\mathbf{f}$ from 2D IQML and $L$ from $R_{\mathbf{nn}}$ the iteration in Equation 173 can be computed. Additionally, since this iteration improves the estimate of $\mathbf{s}$, and consequently the estimate $\mathbf{n}$ at each iteration $\mathbf{s}$ is estimated by Equation 160 and $\mathbf{n}$ is reestimated by Equation 178.

*6.3.2.1 The 2D UCIQML algorithm.* This techniques will be called 2D UCIQML and is summarized as follows:

Step 1. *Estimate the signal* $F\mathbf{s}$. Attain the two dimensional frequencies $(\lambda, \gamma)_i$ for $i = 1 \ldots p$ from 2D IQML. Form the matrix $F$ from the frequencies. Then estimate $\mathbf{s}$ as

$$\mathbf{s} = F^{+}\mathbf{y}. \tag{182}$$

Step 2. *Estimate the noise sequence* $\mathbf{n}$ as

$$\mathbf{n} = \mathbf{y} - F\mathbf{s}. \tag{183}$$

Step 3. *Estimate the autocorrelation of the noise* $r_{NN}$. Form the block Toeplitz matrix $R_{\mathbf{nn}}$ and its Cholesky decomposition $R_{\mathbf{nn}} = L^*L$.

Step 4. *Estimate new* $(\lambda, \gamma)_i$ for $i = 1 \ldots p$. Form $W_{i-1}$ from the last estimate $\mathbf{f}$, $\mathbf{f}_{i-1}$, and iterate once

$$\min_{\mathbf{f}_i} \left\| ((LW_{i-1}W_{i-1}^*L^*)^+ LW_{i-1})\tilde{Y}\mathbf{f}_i \right\|_F^2 \tag{184}$$

Step 5. *Re-estimate* $\mathbf{s}$ as

$$\hat{\mathbf{s}} = ((L^*)^+F)^+(L^*)^+\mathbf{y}. \tag{185}$$

Step 6. *Repeat* Step 2 - 5 for several iterations, or until estimates do not significantly change.

*6.3.3 2-D Exponential Noise.* This section develops a new spectral model for 2-D signals. Algorithms to exploit an AR or ARMA model in two dimensions do not follow as readily as in one dimension since these models do not have the same properties in two dimensions as they do in one dimension. The rational polynomial model for colored noise is less appealing due to the absence of a fundamental theorem of algebra in two dimensions. The 2-D AR model is not efficient on a finite region of support and the set of AR coefficients describing a specific PSD is not unique. The estimated parameters of the rational polynomial noise models in two dimensions vary according to the technique used to estimate the model parameters.

Instead, a suitable low-order noise model with unique parameters is a filter, whose impulse response is the sum of damped exponentials. In 1-D, this model is a subset of the ARMA noise model. The sum of $q$ damped exponentials is synonymous with the 1-D ARMA$(q-1, q)$ model. The one-dimensional ARMA$(q-1, q)$ model transfer function (z-transform domain) in polynomial, pole-zero and partial fraction form may be written as

$$\frac{C(z)}{B(z)} = \frac{\sum_{l=0}^{q-1} c_l z^{-l}}{\sum_{k=0}^{q} b_k z^{-k}} = \frac{\prod_{l=1}^{q-1}(1-\zeta_l z^{-1})}{\prod_{k=1}^{q}(1-\beta_k z^{-1})} = \sum_{k=1}^{q} \frac{s_k}{(1-\beta_k z^{-1})}. \tag{186}$$

The 1-D AR$(q)$ case, excepting the constant term, the coefficients of numerator polynomial in 1-D ARMA$(q-1, q)$ model equal zero and

$$\frac{C(z)}{B(z)} = \frac{1}{\sum_{k=0}^{q} b_k z^{-k}} = \frac{1}{\prod_{k=1}^{q}(1-\beta_k z^{-1})} = \sum_{k=1}^{q} \frac{s_k}{(1-\beta_k z^{-1})}. \tag{187}$$

In both cases, the impulse response of the transfer function $H(z) = C(z)/B(z)$ is sum of $q$ exponentials,

$$h[n] = \sum_{k=1}^{q} s_k \beta_k^n \quad n = 0 \ldots N-1. \tag{188}$$

Thus, the filter for an AR process may be implemented with either the IIR filter coefficients, $b_k$ for $k = 1 \ldots q$, or the FIR filter coefficients, $c_l = \sum_{k=1}^{q} s_k \beta_k^l$ for $l = 0 \ldots N-1$. The exponential amplitudes, $s_k$ for $k = 1 \ldots q$, can be adjusted to create any ARMA$(q-1, q)$ filter.

If this idea is expanded to two dimensions, then, the pole locations of any 2-D filter can be fixed and the 2-D AR coefficients are not needed. Thus the impulse response

$$h[m_1, m_2] = \sum_{k=1}^{q} s_k \beta_k^{m_1} \xi_k^{m_2} \quad 0 \le m_1 \le M_1 - 1 \quad 0 \le m_2 \le M_2 - 1 \tag{189}$$

represents the coefficients of a 2-D filter with known pole locations. This filter is the unique pairing of two 1-D ARMA spectra and suitably constrains our noise model. Equation 189 represents a new 2-D spectral model. For the case where $q = 1$ this defines a 2-D AR filter. In 1-D, a higher order AR filter could be formed by cascading (convolving) several AR(1) filters. In 2-D, however, this results in a filter with poles at all combinations of the spatial frequencies in each dimension. That is, the filter becomes separable and less general. In the separable noise model, the 2-D covariance matrix is independent in each dimension and is the Kronecker product of the covariance matrix for each dimension,

$$R_{\mathbf{nn}} = R_{M_2} \otimes R_{M_1}. \tag{190}$$

Colored noise techniques developed in this chapter will be tested with both the exponential and separable noise model.

*6.3.4 2-D Exponential Noise Maximum Likelihood.* This section develops another new technique to estimate 2-D exponentials in unknown colored noise by extending the 2D IQML technique. With the Exponential noise model, the 2-D extension of the ARUCIQML technique now follows directly. Again the only difference in modeling the exponential noise sequence and exponentials is the input to the filter that produces them, white noise or an impulse. Thus the exponentials and noise poles may be estimated simultaneously by estimating $p + q$ exponentials with the 2D IQML method. To choose which exponentials are signal and which are noise, the

exponential amplitudes are estimated from Equation 179. The $p$ exponentials with the largest amplitude are chosen to represent the signal and the remaining exponentials are chosen as the noise poles. The 2-D autocorrelation matrix, $\hat{R}_{\mathbf{nn}}$, for exponential noise can then be formed. With the estimate of $\mathbf{f}$ from the $p$ 2-D exponentials chosen as signal with 2D IQML, and $L$ from the Cholesky decomposition of $\hat{R}_{\mathbf{nn}}$, the iteration in Equation 173 can be computed.

*6.3.4.1 The 2D ARIQML algorithm.* This algorithm will be called 2D ARIQML because of its similarity to the 1-D method ARUCIQML and is summarized as follows:

Step 1. *Estimate $p$ exponentials and $q$ noise poles and their amplitudes.* Attain the two dimensional frequencies and amplitudes $(\lambda, \gamma, s)_i$ for $i = 1 \ldots p+q$ from 2D IQML

Step 2. *Determine which $(\lambda, \gamma, s)_i$ for $i = 1 \ldots p+q$ are exponentials and which are noise poles.* For high signal-to-noise ratios (SNRs) the $p$ exponentials of with the largest amplitude are the signal exponentials $(\lambda, \gamma, s)_i$ for $i = 1 \ldots p$; the remaining $(\lambda, \gamma, s)_i$ for $i = p+1 \ldots q$ are the noise poles.

Step 3. *Construct $W$ and $L$.* Form $W$ from the $(\lambda, \gamma)_i$ for $i = 1 \ldots p$. Calculate the autocorrelation coefficients of exponential noise poles. Then form the Toeplitz matrix $\hat{R}_{\mathbf{nn}}$ and its Cholesky decomposition $\hat{R}_{\mathbf{nn}} = L^* L$.

Step 4. *Estimate new $(\lambda, \gamma)_i$* for $i = 1 \ldots p$. Iterate once

$$\min_{\mathbf{f}_i} \left\| ((LW_{i-1}W_{i-1}^* L^*)^+ LW_{i-1}) \tilde{Y} \mathbf{f}_i \right\|_F^2 \tag{191}$$

Step 5. Form $W_{i-1}$ from the last estimate of $\mathbf{f}$, $\mathbf{f}_{i-1}$,and repeat Step 4 for several iterations, or until estimates do not significantly change.

Step 6. *Estimate* $\mathbf{s}$ as

$$\hat{\mathbf{s}} = ((L^*)^+ F)^+ (L^*)^+ \mathbf{y}. \tag{192}$$

## *6.4 Stochastic Maximum Likelihood*

This section extends the 2D MODE technique to colored noise and develops two new techniques to estimate 2-D exponentials in unknown colored noise. 2-D stochastic maximum likelihood techniques hinge on an accurate estimate of the 2-D data covariance matrix. For a single data

instance, the outer product estimate of the covariance matrix is only rank one and thus of limited utility. Other estimates of the 2-D covariance matrix such as the outer product of the MEMP data matrix are not full size and thus provide limited accuracy. The assumption of stationarity or of separability leads to a full rank and full size 2-D covariance matrix (Equation 180 and 190, respectively). But, because the 2-D covariance matrix has $2M_1M_2 - M_1 - M_2 + 1$ independent parameters [14], and only $M_1M_2$ samples are available in a single data instance, the full 2-D stationary and separable covariance matrix is of limited use for high resolution estimation of exponentials. The independent components of the covariance matrix in Equation 190, however, provide the basis for computationally less intensive colored noise algorithms based on a 1-D by 1-D formulation.

The 2-D techniques based on 2D IQML in the previous section are computationally intensive, $O(M_1^3M_2^3)$, especially as $M_1$ and $M_2$ increase. A computational savings is attained when colored noise techniques are based the separable model such as the stochastic maximum likelihood technique 2D MODE [38]. The 2D MODE technique minimizes an alternative form of Equation 169

$$\min_{W} \operatorname{tr}\left(P_{L,W} L^{-*} \hat{R}_{\mathbf{yy}} L^{-1}\right) \tag{193}$$

where $P_{L,W}$ is given in Equation 167 and for the single data instance $\hat{R}_{\mathbf{yy}} = \mathbf{yy}^*$. The 2D MODE technique assumes that the exponentials are any possible pairing of the estimated frequencies (a separable model). Thus the matrix $F$ is defined as the Kronecker product of $G$ and $H$,

$$F = H \otimes G. \tag{194}$$

To extend the 2D MODE technique to colored noise we will also assume that the noise is separable as in Equation 190 then

$$\begin{aligned}
P_{L,W} &= I - P_{L,F} \\
&= I - L^{-*}F(F^*R_{\mathbf{nn}}^{-1}F)^{-1}F^*L^{-1} \\
&= I - (L_{M_2}^{-*} \otimes L_{M_1}^{-*})(H \otimes G)((H^* \otimes G^*)(R_{M_2}^{-1} \otimes R_{M_1}^{-1})(H \otimes G))^{-1}(H^* \otimes G^*)(L_{M_2}^{-1} \otimes L_{M_1}^{-1}) \\
&= I - L_{M_2}^{-*}H(H^*R_{M_2}^{-1}H)^{-1}H^*L_{M_2}^{-1} \otimes L_{M_1}^{-*}G(G^*R_{M_1}^{-1}G)^{-1}G^*L_{M_1}^{-1} \\
&= I - P_H \otimes P_G
\end{aligned} \tag{195}$$

where $P_H = L_{M_2}^{-*} H (H^* R_{M_2}^{-1} H)^{-1} H^* L_{M_2}^{-1}$ and $P_G = L_{M_1}^{-*} G (G^* R_{M_1}^{-1} G)^{-1} G^* L_{M_1}^{-1}$. Then expressing $P_H$ and $P_G$ in terms of their orthogonal projections $P_H^\perp$ and $P_G^\perp$.

$$\begin{aligned} P_{L,W} &= I - P_H \otimes P_G \\ &= I - (I - P_H^\perp) \otimes (I - P_G^\perp) \\ &= I - (I - P_H^\perp \otimes I - I \otimes P_G^\perp + P_H^\perp \otimes P_G^\perp) \\ &= P_H^\perp \otimes I + I \otimes P_G^\perp - P_H^\perp \otimes P_G^\perp \end{aligned} \tag{196}$$

where $I$ is an identity matrix appropriately sized for each Kronecker product. Equation 193 is then

$$\min_{\lambda_i, \gamma_j} \operatorname{tr} \left( (P_H^\perp \otimes I + I \otimes P_G^\perp - P_H^\perp \otimes P_G^\perp) \, L^{-*} R L^{-1} \right) \tag{197}$$

where $R = \mathbf{y}\mathbf{y}^*$. 2D MODE considers the Kronecker product $P_H^\perp \otimes P_G^\perp$ to be higher order and it is neglected. Thus, the minimization to find the $\lambda_i$ and the $\gamma_i$ can be performed separately. Effectively, the trace of the blocks of the 2-D covariance matrix are averaged across one dimension. Since both the 2-D covariance matrix estimated by the outer product of the vectorized data, and the 2-D covariance matrix of a stationary noise process have special structure (an outer product, and Toeplitz blocks, respectively), the average across blocks in each dimension can be computed directly without the computationally intensive and memory consuming task of forming the 2-D covariance matrices. The minimization for $\lambda_i$ is

$$\min_{\lambda_i} \operatorname{tr} \left( (I \otimes P_G^\perp) L^{-*} R L^{-1} \right) = \min_{\lambda_i} \operatorname{tr} \left( P_G^\perp L_{M_1}^{-*} R_\lambda L_{M_1}^{-1} \right) \tag{198}$$

where $R_\lambda \equiv \sum_{m_1=0}^{M_1-1} R_{m_1 m_1}$ and $R_{kl} = R(kM_2+1 : (k+1)M_2, lM_2+1 : (l+1)M_2)$ is an $M_2 \times M_2$ submatrix of $R$. The minimization for $\gamma_i$ is

$$\min_{\gamma_j} \operatorname{tr} \left( (P_H^\perp \otimes I) L^{-*} R L^{-1} \right) = \min_{\gamma_j} \operatorname{tr} \left( P_H^\perp L_{M_2}^{-*} R_\gamma L_{M_2}^{-1} \right) \tag{199}$$

where $(R_\gamma)_{kl} \equiv \operatorname{tr}(R_{kl})$. Both minimizations are then accomplished by the 1-D colored noise MODE techniques discussed in the previous chapter. The 1-D techniques CMODE, UCMODE, and ARUCMODE are used almost directly in the 2-D techniques 2D CMODE, 2D UCMODE, and

2D ARMODE. These minimizations then have the form

$$\min_{\mathbf{a}_i} \mathbf{a}_i^* \tilde{U}^* (A_{i-1}^* \hat{R}_{\mathbf{nn}} A_{i-1})^{-1} \tilde{U} \mathbf{a}_i, \tag{200}$$

where $R_M = L_M^* L_M$ and $\tilde{U} = L_M^* [\ \tilde{U}_1 \quad \tilde{U}_2 \quad \cdots \quad \tilde{U}_p\ ]^T$ is defined such that

$$\tilde{U}_i^T \mathbf{a} = A^* \mathbf{u}_i. \tag{201}$$

Here $U = [\ \mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_p\ ]$ contains the $p$ principal eigenvectors of $L_{M_1}^{-*} R_\lambda L_{M_1}^{-1}$ or $L_{M_2}^{-*} R_\gamma L_{M_2}^{-1}$. The components of the noise covariance matrix $L$, $L_{M_1}$ and $L_{M_2}$, are estimated by forming $\hat{R}_{\mathbf{nn}}$ from Equation 180 as in 2D UCIQML and 2D ARIQML, then averaging in the same manner that $L_{M_1}^{-*} R_\lambda L_{M_1}^{-1}$ and $L_{M_2}^{-*} R_\gamma L_{M_2}^{-1}$ are formed from $L^{-*} R L^{-1}$. The harmonics, $\lambda_i$ or $\gamma_j$, are then the roots of the polynomial given by the coefficients in $\mathbf{a}$. Pairing the $\lambda_i$ and $\gamma_j$ is done with Equation 153. The SVD of $L_{M_1}^{-*} R_\lambda L_{M_1}^{-1}$ and $L_{M_2}^{-*} R_\gamma L_{M_2}^{-1}$ dominates the computations making this an $O(M^3)$ technique where $M = M_1 = M_2$.

*6.4.0.2 The 2D UCMODE algorithm.* The stochastic maximum likelihood technique based on unknown colored noise will be called 2D UCMODE and is summarized as follows:

Step 1. *Estimate the signal* $F\mathbf{s}$. Attain the two dimensional frequencies $(\lambda, \gamma)_i$ for $i = 1 \ldots p$ from 2D MODE. Pair the frequencies by least square amplitude as in Equation 153, however, allow no frequency to be used more than once (to attain distinct frequencies). Form the matrix $F$ from the frequencies. Then estimate $\mathbf{s}$ as

$$\mathbf{s} = F^+ \mathbf{y}. \tag{202}$$

Step 2. *Estimate the noise sequence* $\mathbf{n}$ as

$$\mathbf{n} = \mathbf{y} - F\mathbf{s}. \tag{203}$$

Step 3. *Estimate the autocorrelation of the noise* $r_{NN}$. Form the Toeplitz matrices $R_{M_1}$ and $R_{M_2}$, and find their Cholesky decompositions, $L_{M_1}$ and $L_{M_2}$.

Step 4. *Estimate new* $(\lambda,\gamma)_i$ for $i = 1 \ldots p$. Form $A_{i-1}$ from the last estimate $\mathbf{a}$, $\mathbf{a}_{i-1}$, for each dimension

$$\min_{\mathbf{a}_i} \left\| (A_{i-1}^* L_M^*)^+ \tilde{U} \mathbf{a}_i \right\|_F^2 \tag{204}$$

Step 5. *Re-estimate* $\mathbf{s}$ as

$$\hat{\mathbf{s}} = ((L^*)^+ F)^+ (L^*)^+ \mathbf{y}. \tag{205}$$

Step 6. *Repeat* Step 2 - 5 for several iterations, or until estimates do not significantly change.

*6.4.0.3 The 2D ARMODE algorithm.* The stochastic maximum likelihood techniques based on exponential colored noise will be called 2D ARMODE and is summarized as follows:

Step 1. *Estimate p exponentials and q noise poles and their amplitudes.* Attain the two dimensional frequencies and amplitudes $(\lambda,\gamma,s)_i$ for $i = 1 \ldots p+q$ from 2D MODE. Pair the frequencies by least square amplitude, however, allow no frequency to be used more than once (distinct frequencies).

Step 2. *Determine which* $(\lambda,\gamma,s)_i$ for $i = 1 \ldots p+q$ *are exponentials and which are noise poles.* For high signal-to-noise ratios (SNRs) the $p$ exponentials of with the largest amplitude are the signal exponentials $(\lambda,\gamma,s)_i$ for $i = 1 \ldots p$; the remaining $(\lambda,\gamma,s)_i$ for $i = p+1 \ldots q$ are the noise poles.

Step 3. *Construct A and L.* Form $A$ from the $(\lambda,\gamma)_i$ for $i = 1 \ldots p$. Form the autocorrelation coefficients of the exponential noise poles. Form the Toeplitz matrices $R_{M_1}$ and $R_{M_2}$, and find their Cholesky decompositions, $L_{M_1}$ and $L_{M_2}$.

Step 4. *Estimate new* $(\lambda,\gamma)_i$ for $i = 1 \ldots p$. Form $A_{i-1}$ from the last estimate $\mathbf{a}$, $\mathbf{a}_{i-1}$, for each dimension

$$\min_{\mathbf{a}_i} \left\| (A_{i-1}^* L_M^*)^+ \tilde{U} \mathbf{a}_i \right\|_F^2 \tag{206}$$

Step 5. Form $A_{i-1}$ from the last estimate of $\mathbf{a}$, $\mathbf{a}_{i-1}$and repeat Step 4 for several iterations, or until estimates do not significantly change.

Step 6. *Estimate* $\mathbf{s}$ as

$$\hat{\mathbf{s}} = ((L^*)^+ F)^+ (L^*)^+ \mathbf{y}. \tag{207}$$

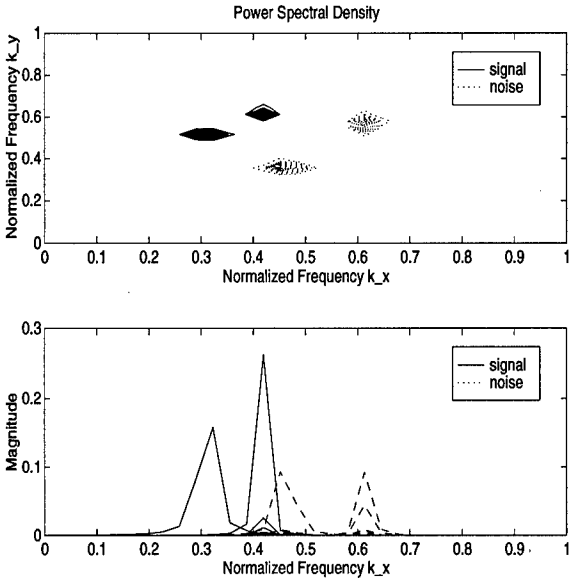


Figure 25. PSD of two exponentials in exponential noise

### *6.5 Results*

Simulations were run with all techniques in this chapter on both the exponential and separable noise models. In the simulations, the 2-D colored noise techniques estimated two modes $(s, \lambda, \gamma)_i = \{(0.9e^{-j0.1\pi}, 0.97e^{-j0.2\pi}, e^{j0.2\pi})_1, (e^{j0.2\pi}, 0.95e^{j0.4\pi}, 1)_2\}$ in exponential noise and separable exponential noise with two frequencies in each dimension $(s_n, \lambda_n, \gamma_n)_i = \{(1, 0.97e^{-j0.1\pi}, 0.97e^{-j0.2\pi})_1, (1, 0.97e^{j0.3\pi}, 0.97e^{j0.1\pi})_2\}$. The PSD of the signal in exponential noise is shown in Figure 25 and the signal in separable noise is shown in Figure 26.

Data instances of $8 \times 8$ samples were used. The results for the deterministic techniques are shown in Figure 27 and 28. The results for the stochastic techniques are shown in Figure 29 and 30. Mean Square Error's (MSE's) were calculated using Monte Carlo simulations each with 200 independent experiments. All of the algorithms were run for three iterations. The straight solid line in the figures is the deterministic 2D CRB which is described in [12].

The performance ($\text{MSE}^{-1}$) of 2D CIQML shows that the 2-D deterministic techniques can attain the CRB. However, no efficient estimator of the 2-D noise was found and the colored noise techniques attain the CRB only asymptotically. The unknown colored noise technique 2D UCIQML, which is not efficient in 1-D (see Chapter 4), and the pole estimating technique 2D ARIQML, which may not be efficient for a finite number of samples in 2-D, did not attain the 2-D CRB [30]. The

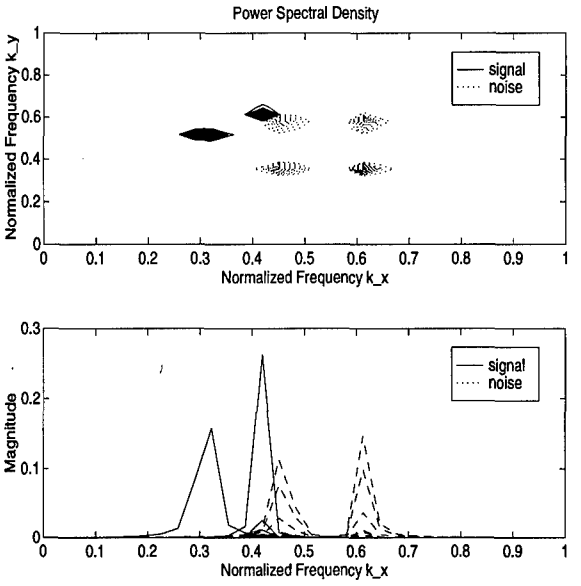


Figure 26. PSD of two exponentials in separable noise

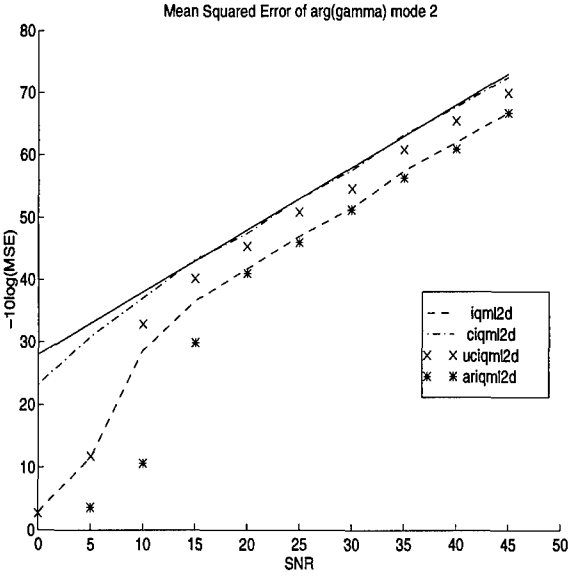


Figure 27. Estimation error for one of two exponentials in exponential noise (Deterministic ML). Techniques: 2-D iterative quadratic maximum likelihood (iqml2d), known colored noise iqml2d (ciqml2d), unknown colored noise iqml2d (uciqml2d), and exponential noise uciqml2d (ariqml2d)

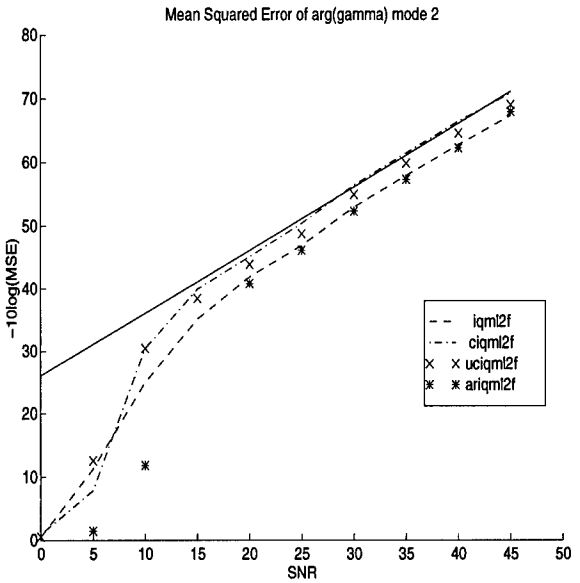


Figure 28. Estimation error for one of two exponentials in separable noise (Deterministic ML). Techniques: 2-D iterative quadratic maximum likelihood, fast version (iqml2f), known colored noise iqml2f (ciqml2f), unknown colored noise iqml2f (uciqml2f), and exponential noise uciqml2f (ariqml2f)

techniques did improve performance over the white noise techniques 2D IQML and 2D MODE. An exception is the 2D ARIQML technique which does not estimate the noise poles accurately enough in comparison to the signal poles. Thus, the inaccurate noise covariance estimate degrades 2D ARIQML estimation performance. This is true despite the fact that in estimates of the noise only 2D ARIQML estimated the noise better than 2D ARMODE, which shows improvement using its noise estimate. The noise estimates of both the AR related techniques were rather weak, identifying noise poles to only within a quadrant of the complex plane. Thus, it is likely the 2D IQML estimates were already accurate enough that no benefit was gained from the weak noise estimates. The performance of the techniques in separable noise was slightly better with respect to the CRB. Of the 2-D unknown noise techniques 2D UCIQML improves over the performance of 2D IQML in colored noise, and 2D UCMODE performs better then 2D ARMODE. The 2D MODE and related colored noise techniques fall short of the full 2D CRB, as they are expected to do, since they are 1-D by 1-D techniques [52].

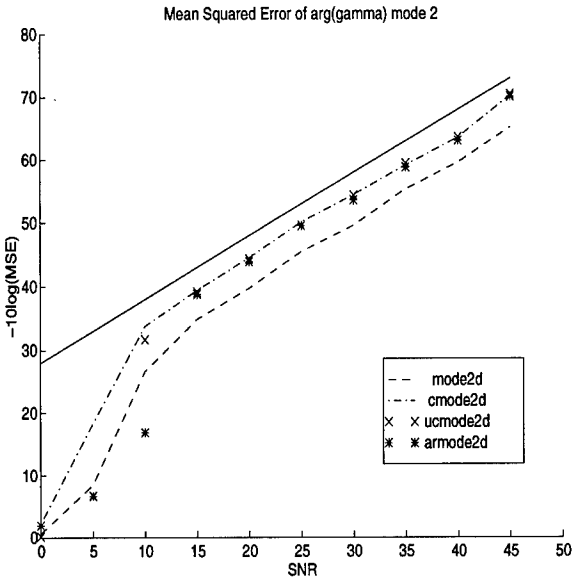


Figure 29. Estimation error for one of two exponentials in exponential noise (Stochastic ML). Techniques: 2-D method of direction estimation (mode2d), known colored noise mode2d (cmode2d), unknown colored noise mode2d (ucmode2d), and exponential noise ucmode2d (armode2d)

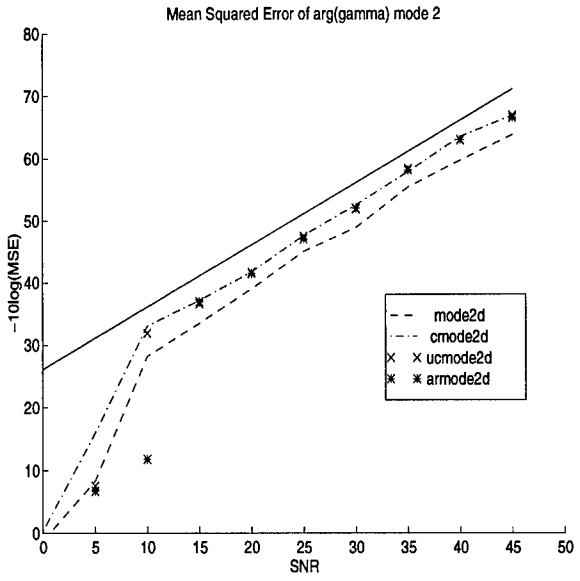


Figure 30. Estimation error for one of two exponentials in separable noise (Stochastic ML). Techniques: 2-D method of direction estimation (mode2d), known colored noise mode2d (cmode2d), unknown colored noise mode2d (ucmode2d), and exponential noise ucmode2d (armode2d)

# *VII. Application to Synthetic Aperture Radar*

In this chapter, the 1-D and 2-D colored noise techniques developed in the previous chapters are applied to simulated and radar chamber SAR data. As a measure of merit, the residual error between the estimated model and the data is used to determine how well the model fits the data. This metric is used to determine if the colored noise algorithms provide better estimates of exponential content of the data than their white noise counterparts. In addition, the scatterer locations identified by the 2-D techniques are examined for suitability as pattern recognition features.

## *7.1 1-D SAR Data*

To test the algorithms on 1-D radar data, two down-range profiles (cross-range data was less well modeled as damped exponentials) were selected from the SAR image of a tank produced by XPATCH software [1]. The profiles are shown in Figures 31 and 34. A damped exponential model was fit to each profile with colored and white noise algorithms for model orders $p = 1 \ldots 10$. A noise model order of $q = p$, or $q = 2$ was used for the unknown colored noise algorithms. The normalized error between the estimated model, $\hat{\mathbf{y}}$, and the radar profile $\mathbf{y}$, was calculated as

$$\bar{e} = \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|_2}{\|\mathbf{y}\|_2}$$

where $\|\mathbf{y}\|_2 = \sqrt{\mathbf{y}^*\mathbf{y}}$. The resulting modeling error is shown in Figures 32 through 36. The results show that for low model order $p$, the ARUCIQML and ARMAIQML algorithms fit the data better. As the model order increases, IQML and UCIQML continually model more of the exponential signal and noise, and better fit the data. Use of ARUCIQML and ARMAIQML fits this radar data better than IQML at low model orders and may provide mode accurate relative scatterer locations (exponential frequencies) for target identification in this case. Results for the stochastic algorithm for a noise model order of $q = p$ are shown in Figures 37 and 38.

As was indicated in the development of the 1-D colored noise algorithms, there is an easy way to increase the accuracy of white noise algorithms. Simply fit a larger model order to the data than the expected number of signals. Then, pick the highest energy part of the estimate as signal. This is called overmodeling. It is also the initial step in the AR and ARMA colored noise algorithms. Including the overmodeling technique, IQMLO, the resulting modeling error is shown in Figures 39 and 40. A noise model order of $q = p$ was used for the unknown colored noise algorithms.

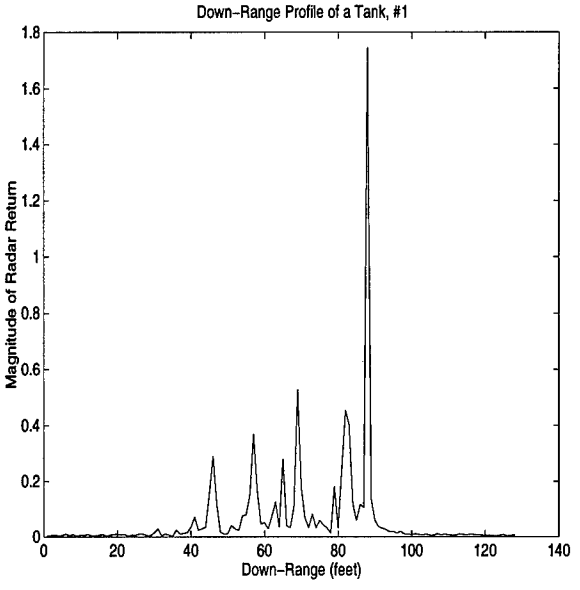


Figure 31. Down-range radar profile 1

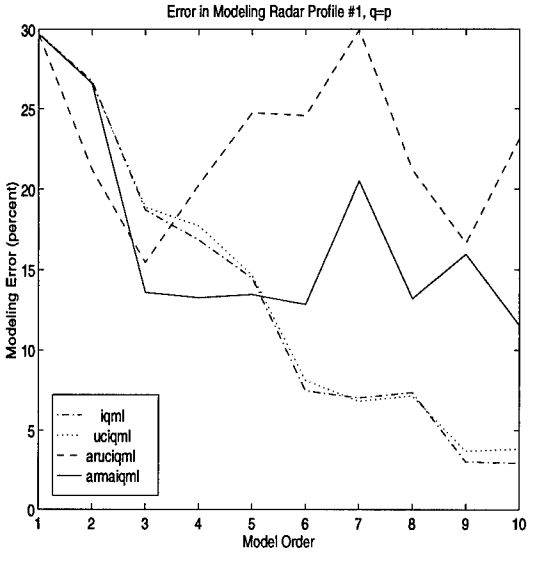


Figure 32. Modeling error using $p$ noise poles (Deterministic ML, Profile 1). Techniques: iterative quadratic maximum likelihood (iqml), unknown colored noise iqml (uciqml), autoregressive uciqml (aruciqml), and autoregressive moving average uciqml (armaiqml)

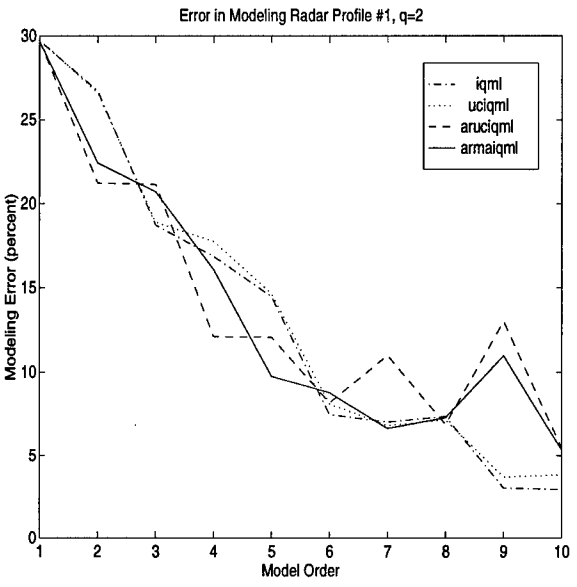


Figure 33. Modeling error using two noise poles (Deterministic ML, Profile 1). Techniques: iterative quadratic maximum likelihood (iqml), unknown colored noise iqml (uciqml), autoregressive uciqml (aruciqml), and autoregressive moving average uciqml (armaiqml)

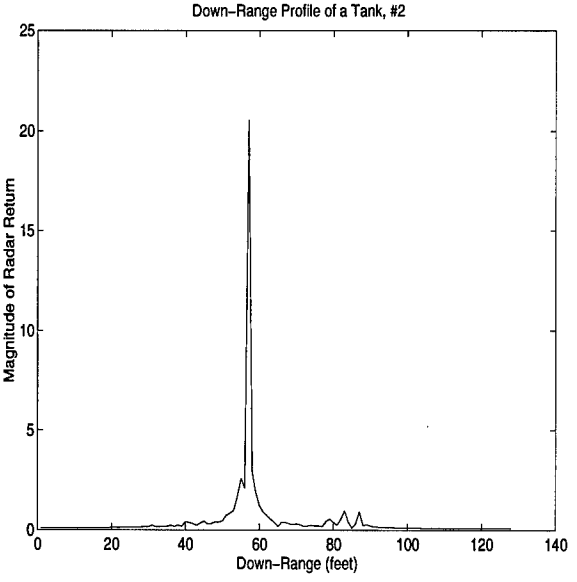


Figure 34. Down-range radar profile 2

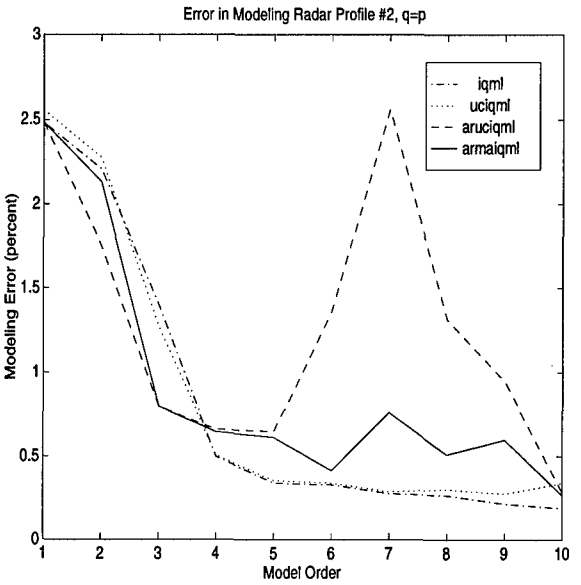


Figure 35. Modeling error using $p$ noise poles (Deterministic ML, Profile 2). Techniques: iterative quadratic maximum likelihood (iqml), unknown colored noise iqml (uciqml), autoregressive uciqml (aruciqml), and autoregressive moving average uciqml (armaiqml)

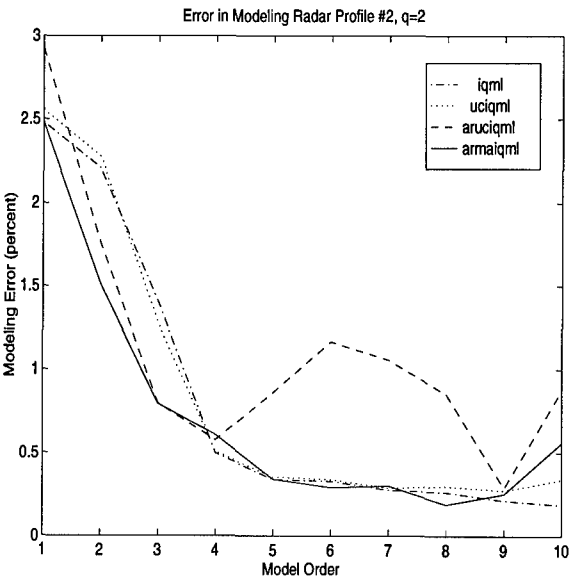


Figure 36. Modeling error using two noise poles (Deterministic ML, Profile 2). Techniques: iterative quadratic maximum likelihood (iqml), unknown colored noise iqml (uciqml), autoregressive uciqml (aruciqml), and autoregressive moving average uciqml (armaiqml)

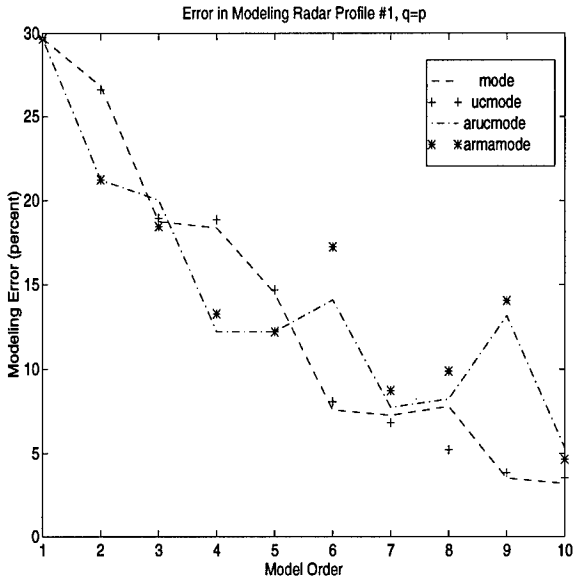


Figure 37. Modeling error using $p$ noise poles(Stochastic ML, Profile 1). Techniques: method of direction estimation (mode), unknown colored noise mode (ucmode), autoregressive ucmode (arucmode), and autoregressive moving average ucmode (armamode)

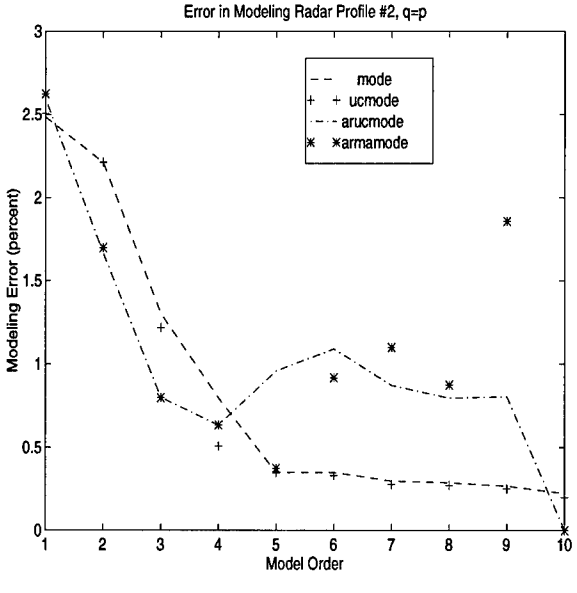


Figure 38. Modeling error using $p$ noise poles(Stochastic ML, Profile 2). Techniques: method of direction estimation (mode), unknown colored noise mode (ucmode), autoregressive ucmode (arucmode), and autoregressive moving average ucmode (armamode)

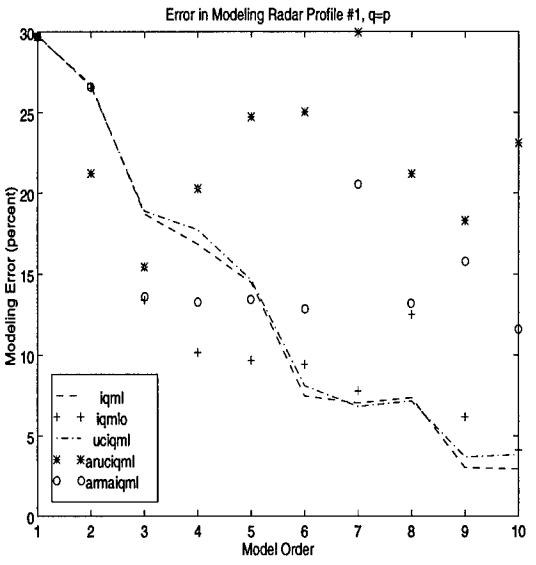


Figure 39. Modeling error using $p$ noise poles (Overmodeling, Profile 1). Techniques: iterative quadratic maximum likelihood (iqml), iqml with p extra noise poles (iqmlo), unknown colored noise iqml (uciqml), autoregressive uciqml (aruciqml), and autoregressive moving average uciqml (armaiqml)

When overmodeling is compared to the colored noise techniques the performance difference is less significant. In this sense, the colored noise algorithms may simply be correcting for a bad choice of model order. From an algorithmic point of view, the difference between overmodeling and the colored noise approaches is clear. Both approaches estimate the noise. Overmodeling discards the noise part of the estimate. Conversely, the colored noise techniques use the noise estimates to improve the signal estimates.

### *7.2 2-D SAR Data*

For 2-D data, bounds on the expected performance of any technique that produces good estimates are described below. These bounds assume that the data are well modeled as a finite sum of damped exponentials. For the Synthetic Aperture Radar (SAR) problem, the $M_1 \times M_2$ focused image, $X$, and interpolated phase history, $W$, are represented as a discrete Fourier transform (DFT) pair by

$$X = D_{M_1} W D_{M_2}^T, \tag{208}$$

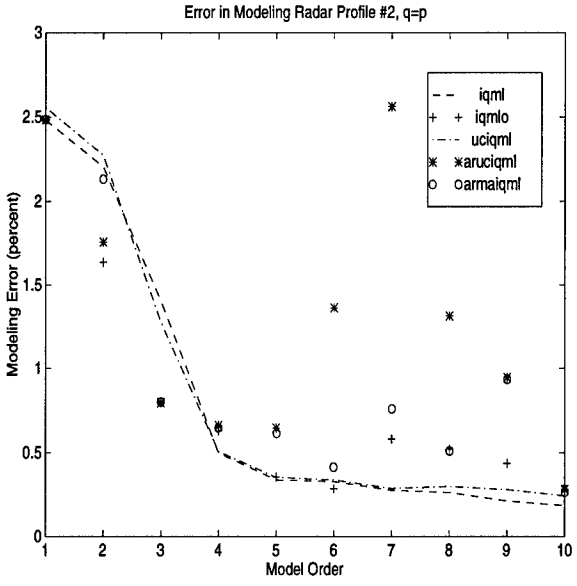


Figure 40. Modeling error using $p$ noise poles (Overmodeling, Profile 2). Techniques: iterative quadratic maximum likelihood (iqml), iqml with p extra noise poles (iqmlo), unknown colored noise iqml (uciqml), autoregressive uciqml (aruciqml), and autoregressive moving average uciqml (armaiqml)

where $D_M$ is the normalized DFT matrix with $(k,l)^{th}$ element $e^{-j\frac{2\pi}{M}kl}/\sqrt{M}$, $0 \leq k,l \leq M-1$. A measure of the error between the phase histories, $W$, and the estimated model, $\hat{W}$, is given by

$$e \equiv ||E||_F \equiv ||W - \hat{W}||_F. \tag{209}$$

To bound the performance of parameter estimation schemes one might consider low rank models other than the damped exponential model. A general low rank decomposition of the $M_1 \times M_2$ matrix, $W$, is

$$W = W_p + E, \tag{210}$$

where $W_p$ is rank-$p$ and $E$ is a residual matrix. The best approximation of order $p$ is found by using the SVD of $W$ : $W = U\Sigma V^*$. Here the approximated matrix, $\hat{W}$, can be assembled from rank-$p$ versions of the SVD component matrices as

$$\hat{W} = U(1:p,:)\Sigma(1:p,1:p)V(1:p,:)^*, \tag{211}$$

where $\Sigma$ contains the $p$ largest singular values, the $p$ columns of $U$ are the associated left singular vectors, and the $p$ columns of $V^*$ are the associated right singular vectors. The rank-$p$ SVD approximation provides a lower bound on the error of all other rank-$p$ approximations in the form $W = GSH^T$. The error attained by the rank-$p$ SVD approximation is given by

$$e_{SVD} = \|E_{SVD}\|_F = \sqrt{\sum_{i=p+1}^{\min(M_1,M_2)} \sigma_i^2}, \tag{212}$$

where $\sigma_i$ is the $i^{th}$ singular value in $\Sigma$. Now, consider the rank-$p$ model of 2-D damped exponentials. The columns of $G$ and $H$ are Vandermonde. Since this is a restriction, this limits the accuracy of our rank-$p$ approximation. The error incurred by this assumption is less than that of Equation 212. Thus $e_{damp} \geq e_{SVD}$. Now, consider a further restriction on the structure of $G$ and $H$ where the exponential frequencies $\lambda_i$, and $\gamma_i$ can take on only discrete values, specifically, the Discrete Fourier Transform (DFT) frequencies, $\lambda_i \in \{e^{j\frac{2\pi}{M_1}k}\}_{k=0}^{M_1-1}$ and $\gamma_i \in \{e^{j\frac{2\pi}{M_2}l}\}_{l=0}^{M_2-1}$. Observe from Equations 208 and 209 that the minimum is attained by selecting the spatial frequencies of the $p$ values of $X$ with largest moduli. This DFT frequency assumption represents a restriction to the damped exponential model. The corresponding errors then satisfy $e_{DFT} \geq e_{damp} \geq e_{SVD}$.

*7.2.1 Model Fitting.* The 2-D exponential estimation algorithms were tested with two sources of SAR data whose images are shown in Figures 6, 7, 41, and 42. The first source is the XPATCH radar cross section (RCS) modeling tool. XPATCH computes polar-coordinate RCS measurements from a Computer Aided Design (CAD) model of a target. The data selected is of a generic tank seen from above at a steep look-angle. The data was processed with the SAR simulation software provided with XPATCH, which produced images interpolated to a rectangular grid. Since the entire extent of the data is interpolated, the phase history contains regions of zero measurement. The windowing effect of including these zero measurements in the data degrades exponential estimation. Thus, a non-zero square segment of the phase history data was selected.

The second source of SAR data is radar test chamber data of a scale model of a C-29 aircraft. Algorithms were tested on this unfocused data and on the data focused by two of the methods discussed in Chapter 4. Due to computational considerations, only scattering center focus could be applied to the full $128 \times 128$ data size. Also, due to computational considerations, the techniques based on 2D IQML could only be run on a $16 \times 16$ data size (a comparable computational load to

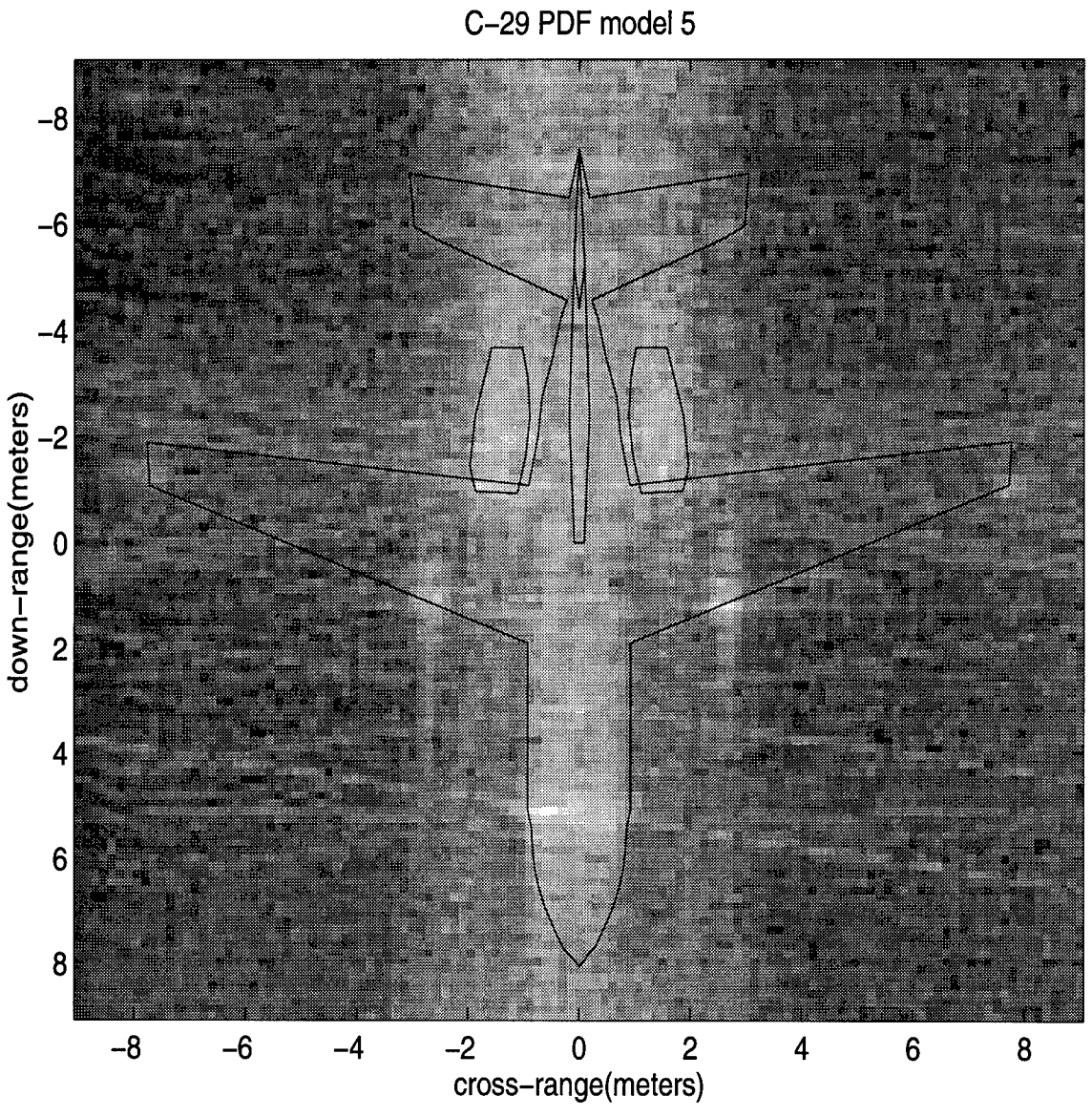


Figure 41. C-29 focused with method from [42] using $500 \times 500$ samples

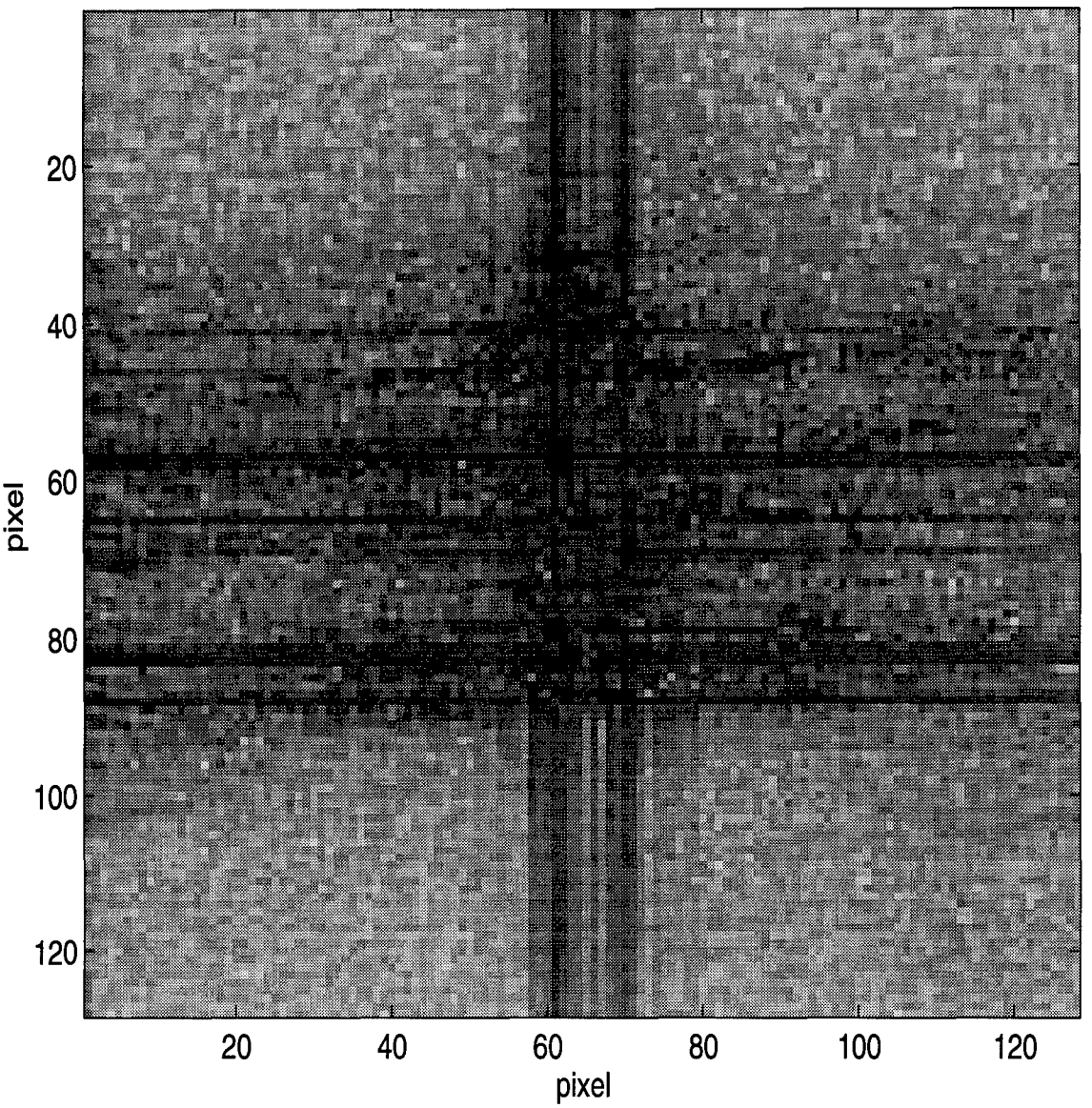


Figure 42. XPATCH tank

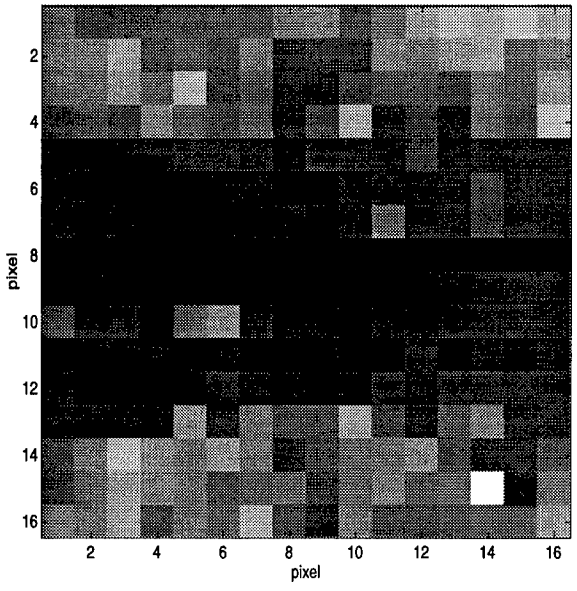


Figure 43. XPATCH tank, $16 \times 16$

running the techniques based on 2D MODE on the $128 \times 128$ data size). Because of the better than Fourier resolution, the scatterer locations of the 2-D techniques are compared to (overlaid on) an image of the full $128 \times 128$ data size. Two method were used to create the focused $16 \times 16$ data. In one case, the $128 \times 128$ scattering center focus image was decimated by 8, and, in the other case, the inverse distance method was used on a $16 \times 16$ sample of the chamber data. The $16 \times 16$ data samples for the XPATCH tank and C-29 aircraft are shown in Figures 43 through 45. An example of the Fourier resolution for $16 \times 16$ data samples (zero-padded in the phase history to $128 \times 128$ samples) is shown in Figure 46 for the XPATCH tank.

For this set of SAR data, the fit of the low rank approximations discussed in the previous section was determined. The normalized error between an estimated model and the phase history was calculated as

$$\bar{e} = \frac{\|W - \hat{W}\|_F}{\|W\|_F}. \tag{213}$$

The error bounds given by the SVD and the DFT were established for this data. These represent lower and upper bounds, respectively, between which any reasonable approximation method should perform. The 2-D white and colored noise techniques developed in the previous chapter were tested with this data. Results for the images are plotted for the deterministic techniques in Figures 47 through 50, for the stochastic techniques in Figures 51 through 53 and 56 through 58 , and for other

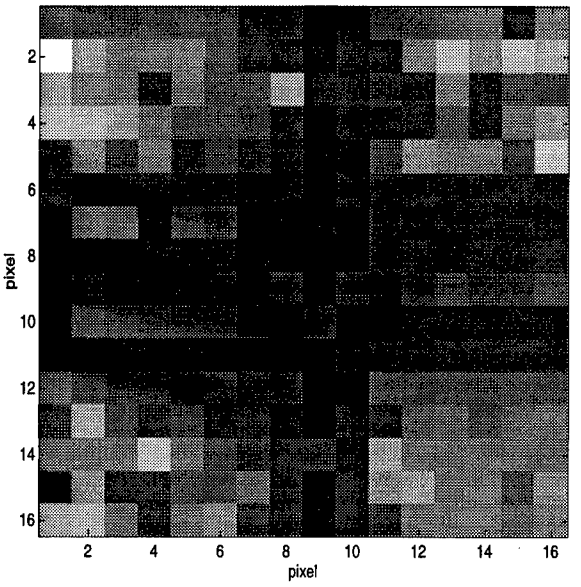


Figure 44. C-29 aircraft, scatter center focus

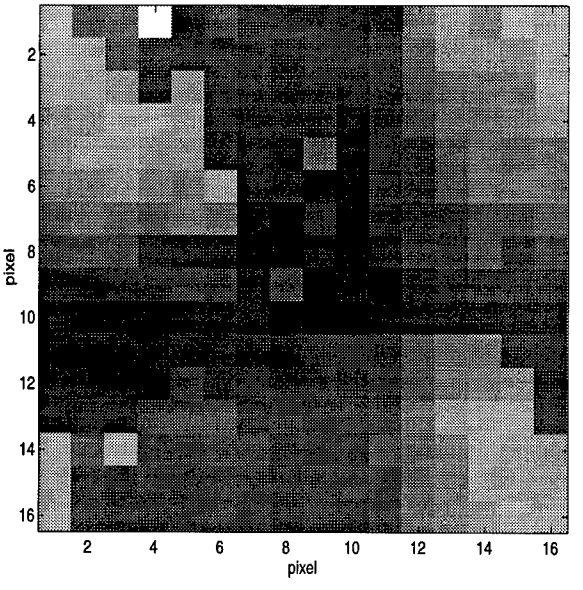


Figure 45. C-29 aircraft, inverse distance focus

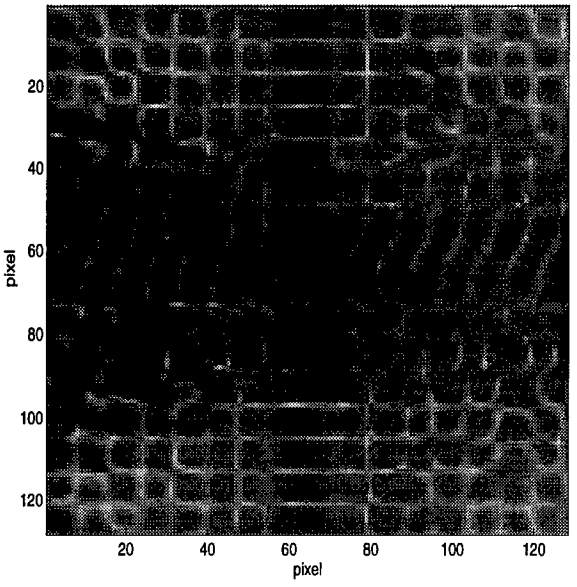


Figure 46. XPATCH tank, $16 \times 16$ zero-padded to $128 \times 128$

2-D techniques in Figure 54, 55, 59 and 60. These results compare how well the exponential content chosen by each technique fits the data. A better fit will result in less error. The actual pole locations are shown later for some sample model orders to understand the suitability of the techniques as region of interest operators, or features for model-based pattern recognition. Locations of the scatterers in the images are plotted for the deterministic techniques in Figures 61 through 64, for the stochastic techniques in Figures 65 through 67 and 70 through 72, and for other 2-D techniques in Figures 68, 69, 73 and 74.

The pairing necessary in the 1-D by 1-D techniques may be accomplished in several ways, depending on an assumption of exponentials with distinct or non-distinct frequencies. The 1-D by 1-D techniques developed in Chapter 6 use distinct exponentials for an initial estimate, then allow the final result to have non-distinct exponentials. This combination results in the best performance in terms of the CRB and model fitting error. A drawback of this combination, the tendency of the 1-D by 1-D methods to pick false scatterers in the same column or row as true scatterers (Figure 72), can be mitigated by other combinations. Other choices of distinct and non-distinct scatters can lead to the identification of additional scattering centers as shown in Figure 75 and 76. One drawback is these cases is in the fully distinct pairing method; erroneous combinations of exponentials will result from the forced pairing. One possible remedy of the pairing problem is to overmodel the

exponentials as in the 2D ARMODE technique then choose the signal exponentials with some full 2-D matching scheme. A similar approach is taken in the 2D Prony matching technique [59].

The matching scheme of 2D Prony estimates the amplitudes of each row (column) based on the exponential frequency estimates for the columns (rows). The amplitudes are then used as data to attain $p$ row exponential frequency estimates to pair with the column (row) frequency. The pairings of exponentials closest to the pairings generated when the roles of row and column are reversed (in parentheses) are matched. A matching technique for the 1-D by 1-D algorithms along the lines of the 2D Prony technique should at least duplicate the good performance of this technique in selecting many individual scatterers and not simply overmodeling high-energy scatterers (Figures 73 and 74). This is indeed the case as shown in Figures 77 and 78. The fit of the data using the 1-D by 1-D stochastic ML techniques is also improved by matching in this manner (Figures 79 and 80). The scattering center locations identified by the 2-D deterministic ML techniques are also improved by performing the estimation twice and matching the results as in 2D Prony (Figures 81 through 84).

### *7.3 Discussion*

Although the data is limited some observations can be made from the results. The error in fitting the data for the 2-D techniques do not in general indicate better performance at low model as the 1-D results do. This is probably due to an inability to model the scatterers well in both dimensions with damped exponentials. The cross-range dimension created by the synthetic aperture is not the smooth damped exponential data seen in the down-range data. Overall, the colored noise techniques show no significant improvement in fitting the data over the white noise techniques.

In the area of focusing, the fit error of the focused C-29 data (Figure 58) vs. the unfocused C-29 data (Figure 57) and the scatterers identified (Figures 72 vs. 71) indicates the improvement gained from focusing. Additionally, the inverse distance focusing (Figures 49 and 52) does work as well with this data as subsampling the scattering center focus data (Figure 50 and 53) does. Scattering center focus is also especially beneficial for the 1-D by 1-D techniques since it is the only computationally tractable focusing method for the full $128 \times 128$ data.

Comparing the fit of different techniques for the same size data set (Figures 47, 51 and 54 and Figures 50, 53 and 55 ) and the scatterer locations identified (Figures 61, 65 and 68 and Figures 64, 67 and 69 ) shows that the full 2-D techniques based on 2D IQML fit the data the best, the 1-D by 1-D techniques based on 2D MODE performed next best, while the remaining 2-D techniques did not fit the data as well.

The high resolution of these techniques is evident in the results. Notice that for most techniques the major scatterers in the $128 \times 128$ image are accurately located despite use of only $16 \times 16$ data samples. For the 1-D by 1-D techniques, however, fewer false scattering locations were identified with the $128 \times 128$ data.

Consistent identification of the same scatterers is important for target recognition. The 2-D techniques based on 2D IQML consistently identify the same scattering centers in one image at different model orders (Figure 61). However, since 2D IQML requires exponentials with distinct frequencies in one dimension, not all the same scatterers are identified when the image is rotated 90 degrees (Figure 62) and a different fit of the model and data is observed (Figures 47 and 48). Scatterer locations are not distinct in both dimensions, thus, they are not correctly located by these techniques in all cases. For this SAR data then, the use of 2D Prony matching with the exponential estimation techniques gives better scatterer locations (Figures 85 and 86).

The colored noise techniques with 2D Prony matching produce a better fit to the data than the white noise techniques at many model orders. In addition, the colored noise techniques do identify different scattering centers than the white noise techniques. The 1-D by 1-D techniques, as indicated earlier, tend to identify scatterers on a line. Additionally, on the $128 \times 128$ data the colored techniques tended to pick out new scattering centers instead of overmodeling high energy ones. The scattering centers identified, however, are dependent on assumptions of distinct and non-distinct scattering centers, and many false scattering centers are identified as discussed in the previous section. This is then the reverse of the problem of 2D IQML where too few scattering centers are identified because of the technique's requirement for distinct frequencies in one of the dimensions. The solution to this problem is provided by the matching technique in the 2D Prony algorithm. With this matching technique most scattering centers are identified and false scattering centers off the target are eliminated.

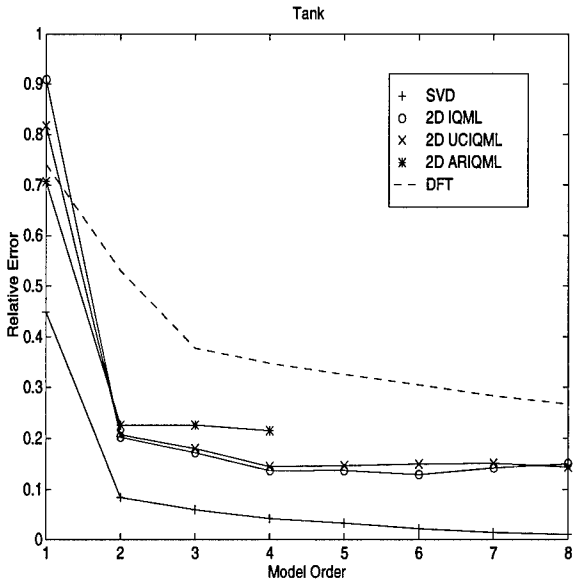


Figure 47. Representation error vs. model order (Deterministic ML, $16 \times 16$ tank image). Techniques: 2-D iterative quadratic maximum likelihood (2D IQML), unknown colored noise iqml2d (2D UCIQML), and exponential noise uciqml2d (2D ARIQML), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT). 2D IQML related techniques are limited to estimating 8 total exponentials and noise poles (4 exponentials and 4 noise poles in 2D ARIQML here)

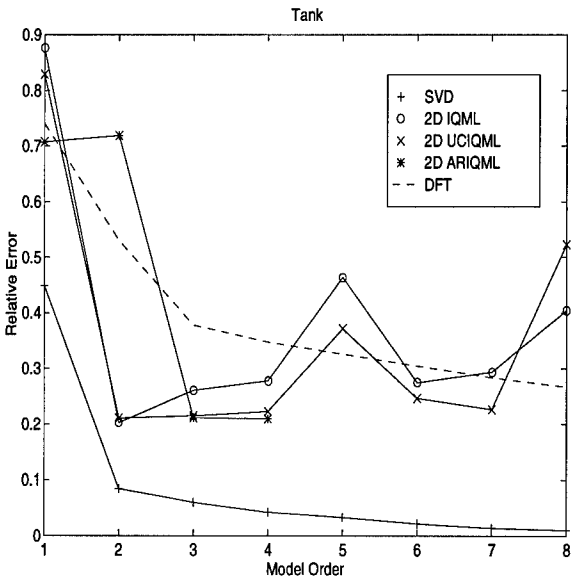


Figure 48. Representation error vs. model order (Deterministic ML, $16 \times 16$ rotated tank image). Techniques: 2-D iterative quadratic maximum likelihood (2D IQML), unknown colored noise iqml2d (2D UCIQML), and exponential noise uciqml2d (2D ARIQML), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

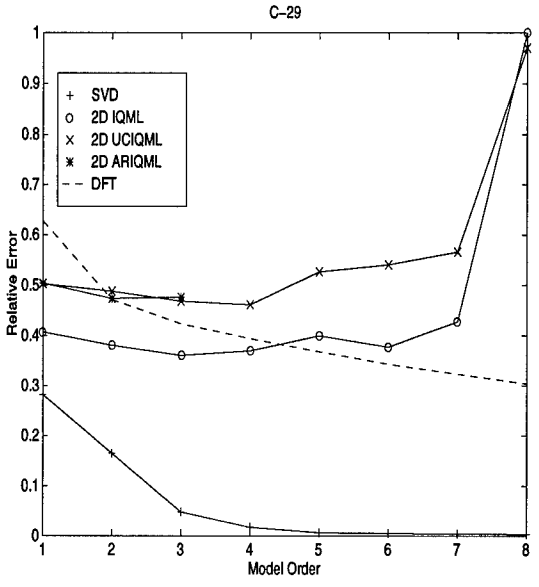


Figure 49. Representation error vs. model order (Deterministic ML, $16 \times 16$ C-29 inverse distance focus image). Techniques: 2-D iterative quadratic maximum likelihood (2D IQML), unknown colored noise iqml2d (2D UCIQML), and exponential noise uciqml2d (2D ARIQML), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

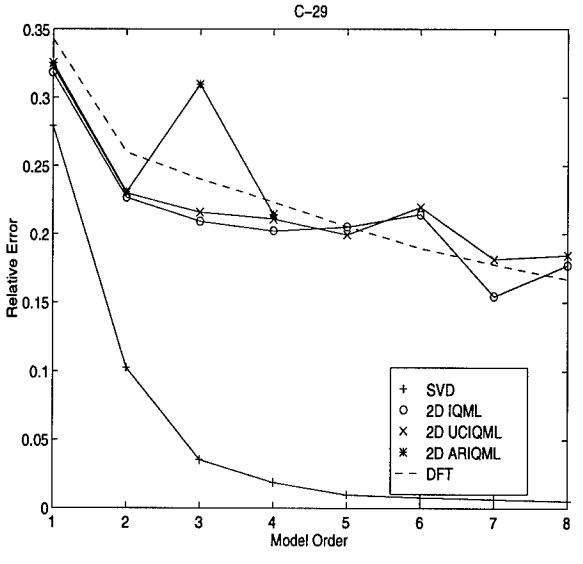


Figure 50. Representation error vs. model order (Deterministic ML, $16 \times 16$ C-29 scattering center focus image). Techniuqes: 2-D iterative quadratic maximum likelihood (2D IQML), unknown colored noise iqml2d (2D UCIQML), and exponential noise uciqml2d (2D ARIQML), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

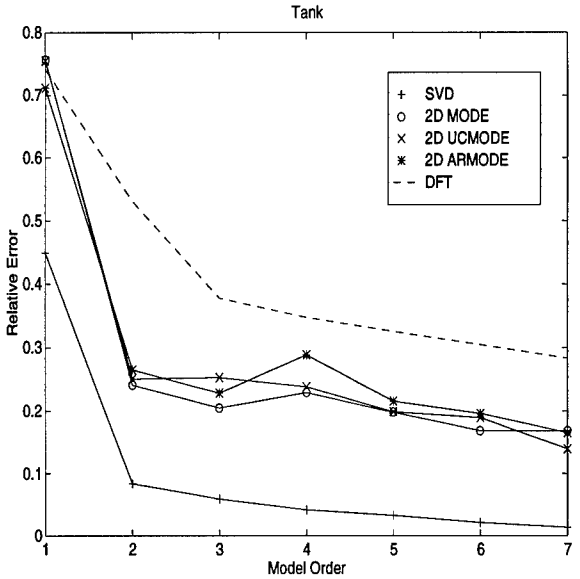


Figure 51. Representation error vs. model order (Stochastic ML, 16×16 tank image). Techniques: 2-D method of direction estimation (2D MODE), unknown colored noise mode2d (2D UCMODE), and exponential noise ucmode2d (2D ARMODE), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

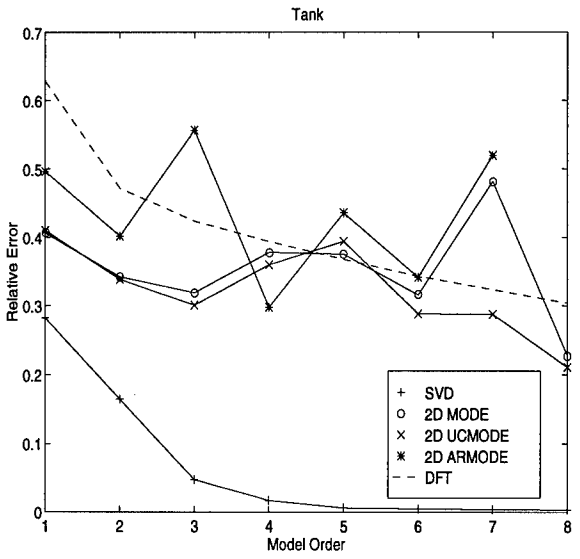


Figure 52. Representation error vs. model order (Stochastic ML, 16 × 16 C-29 inverse distance focus image). Techniques: 2-D method of direction estimation (2D MODE), unknown colored noise mode2d (2D UCMODE), and exponential noise ucmode2d (2D ARMODE), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

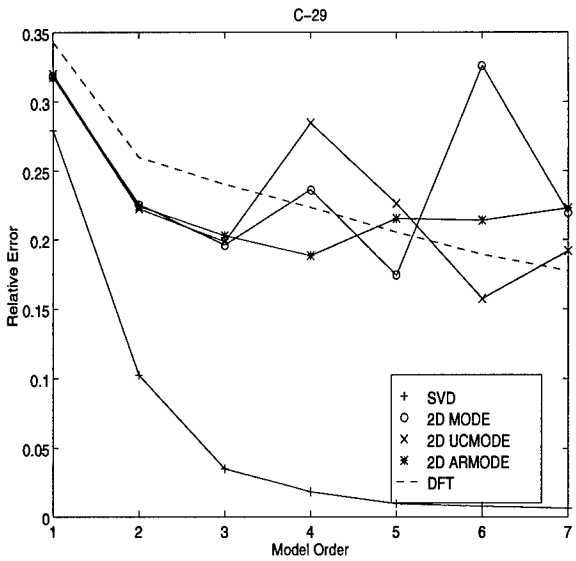


Figure 53. Representation error vs. model order (16 × 16 C-29 scattering center focus image). Techniques: 2-D method of direction estimation (2D MODE), unknown colored noise mode2d (2D UCMODE), and exponential noise ucmode2d (2D ARMODE), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

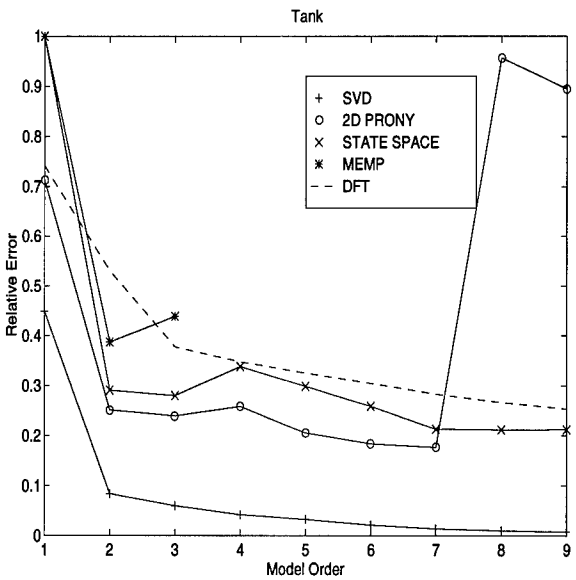


Figure 54. Representation error vs. model order (Other techniques, 16 × 16 tank image). Techniques: 2-D Prony (2D PRONY), state space (STATESP), and matrix enhancement matrix pencil (MEMP) methods, Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

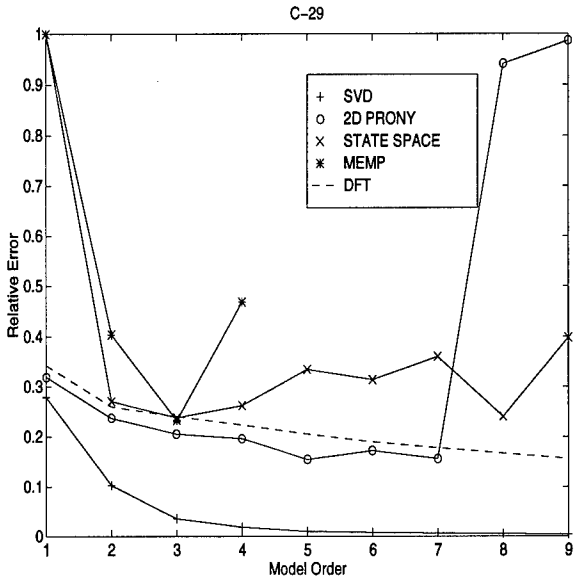


Figure 55. Representation error vs. model order (Other techniques, $16 \times 16$ C-29 scattering center focused image). Techniques: 2-D Prony (2D PRONY), state space (STATESP), and matrix enhancement matrix pencil (MEMP) methods, Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

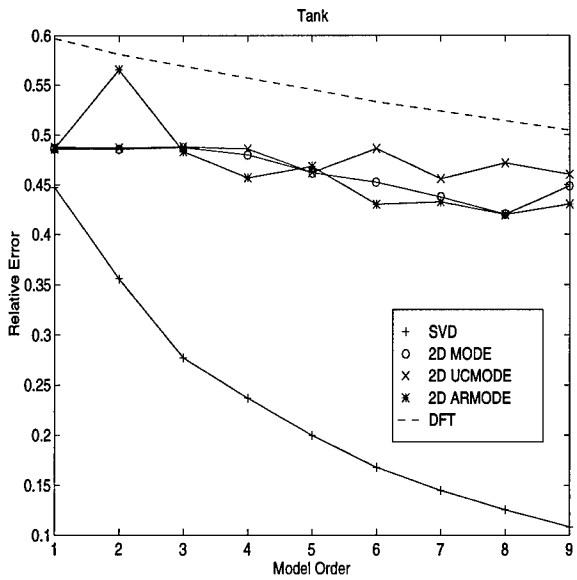


Figure 56. Representation error vs. model order (Stochastic ML, $128 \times 128$ tank image). Techniques: 2-D method of direction estimation (2D MODE), unknown colored noise mode2d (2D UCMODE), and exponential noise ucmode2d (2D ARMODE), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

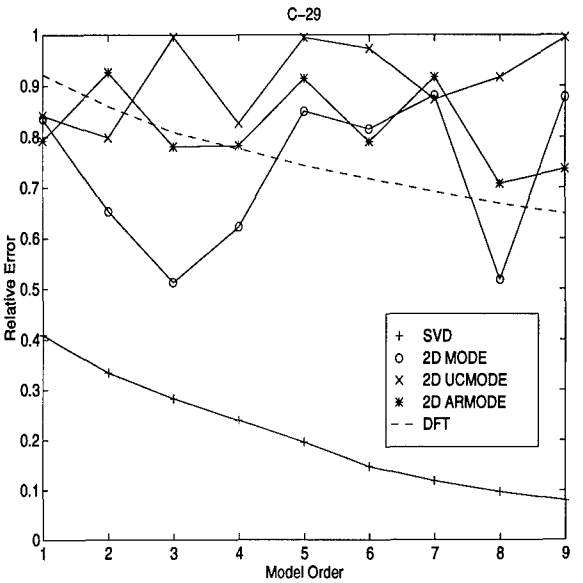


Figure 57. Representation error vs. model order (Stochastic ML, 128×128 C-29 unfocused image). Techniques: 2-D method of direction estimation (2D MODE), unknown colored noise mode2d (2D UCMODE), and exponential noise ucmode2d (2D ARMODE), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

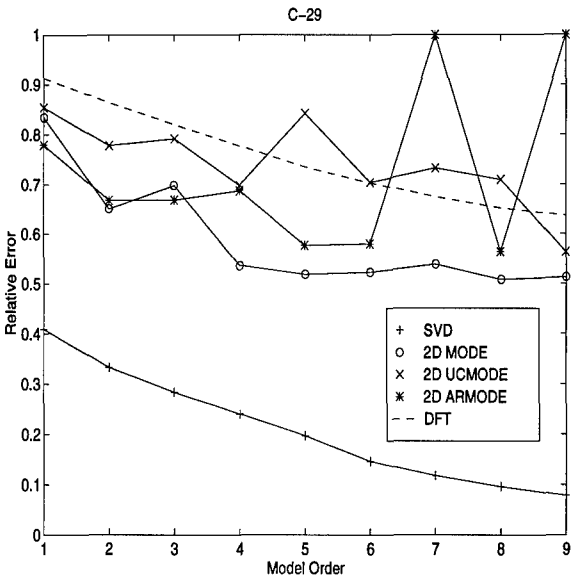


Figure 58. Representation error vs. model order (Stochastic ML, 128 × 128 C-29 scattering center focused image). Techniques: 2-D method of direction estimation (2D MODE), unknown colored noise mode2d (2D UCMODE), and exponential noise ucmode2d (2D ARMODE), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

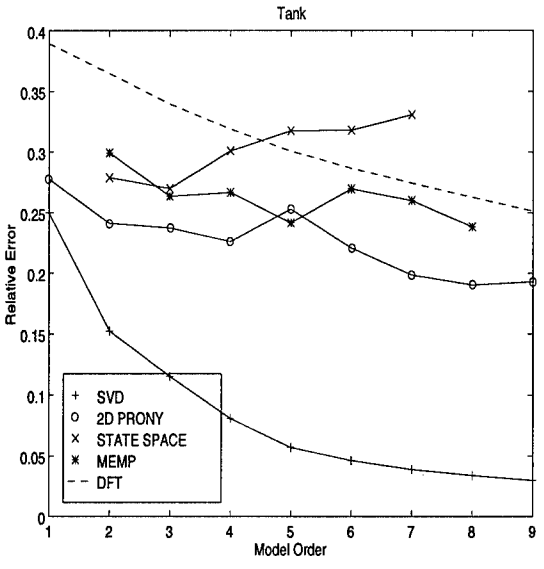


Figure 59. Representation error vs. model order (Other techniques, $64 \times 64$ tank image). Techniques: 2-D Prony (2D PRONY), state space (STATESP), and matrix enhancement matrix pencil (MEMP) methods, Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

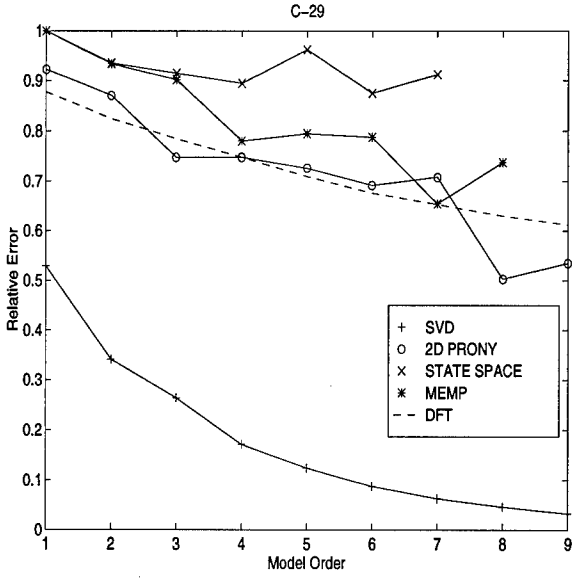


Figure 60. Representation error vs. model order (Other techniques, $64 \times 64$ C-29 scattering center focused image). Techniques: 2-D Prony (2D PRONY), state space (STATESP), and matrix enhancement matrix pencil (MEMP) methods, Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

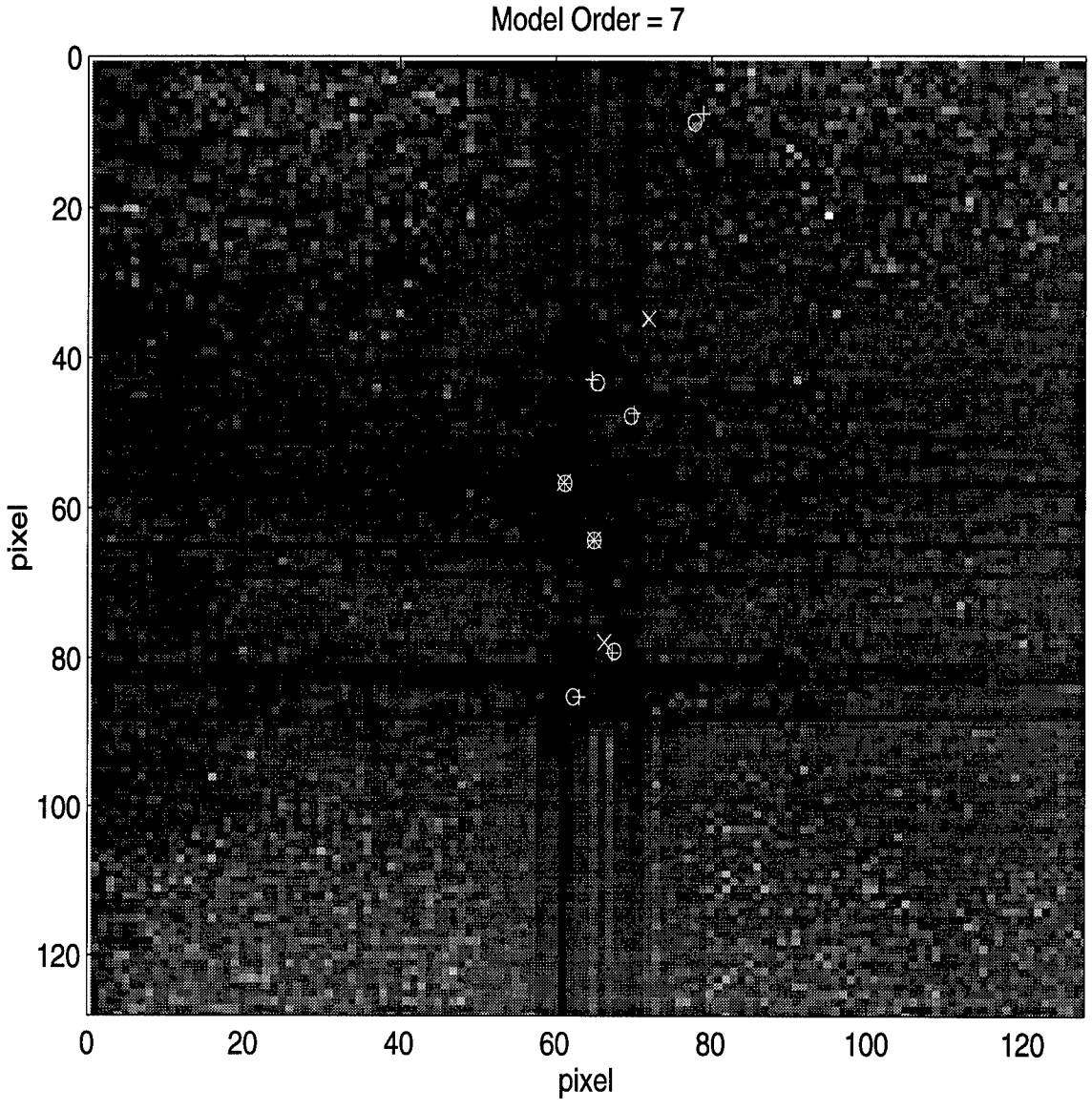


Figure 61. Deterministic ML scatterers (16x16 tank image).
o = 2D IQML + = 2D UCIQML x = 2D ARIQML

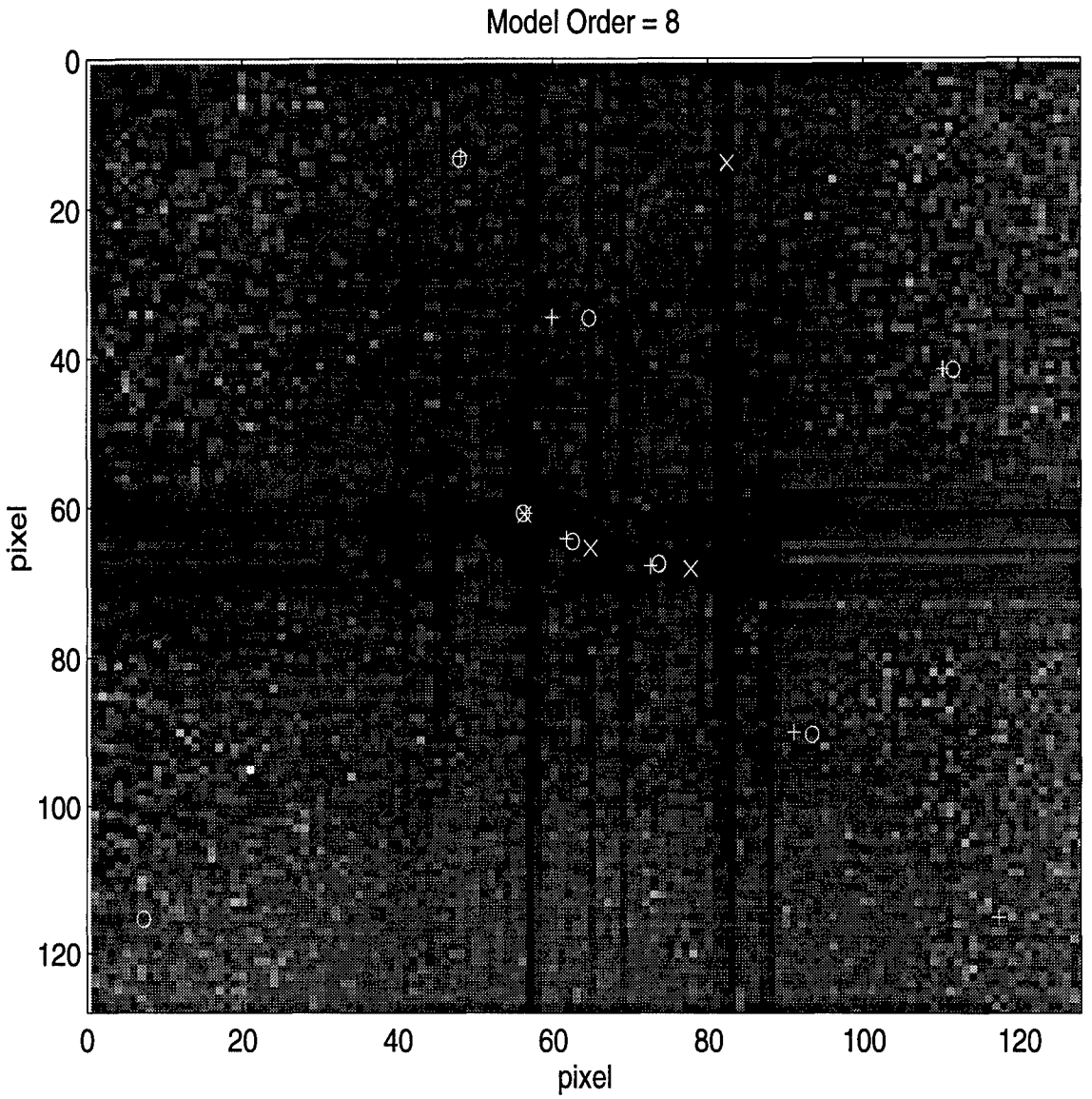


Figure 62. Deterministic ML scatterers (16 × 16 rotated tank image).
o = 2D IQML + = 2D UCIQML x = 2D ARIQML

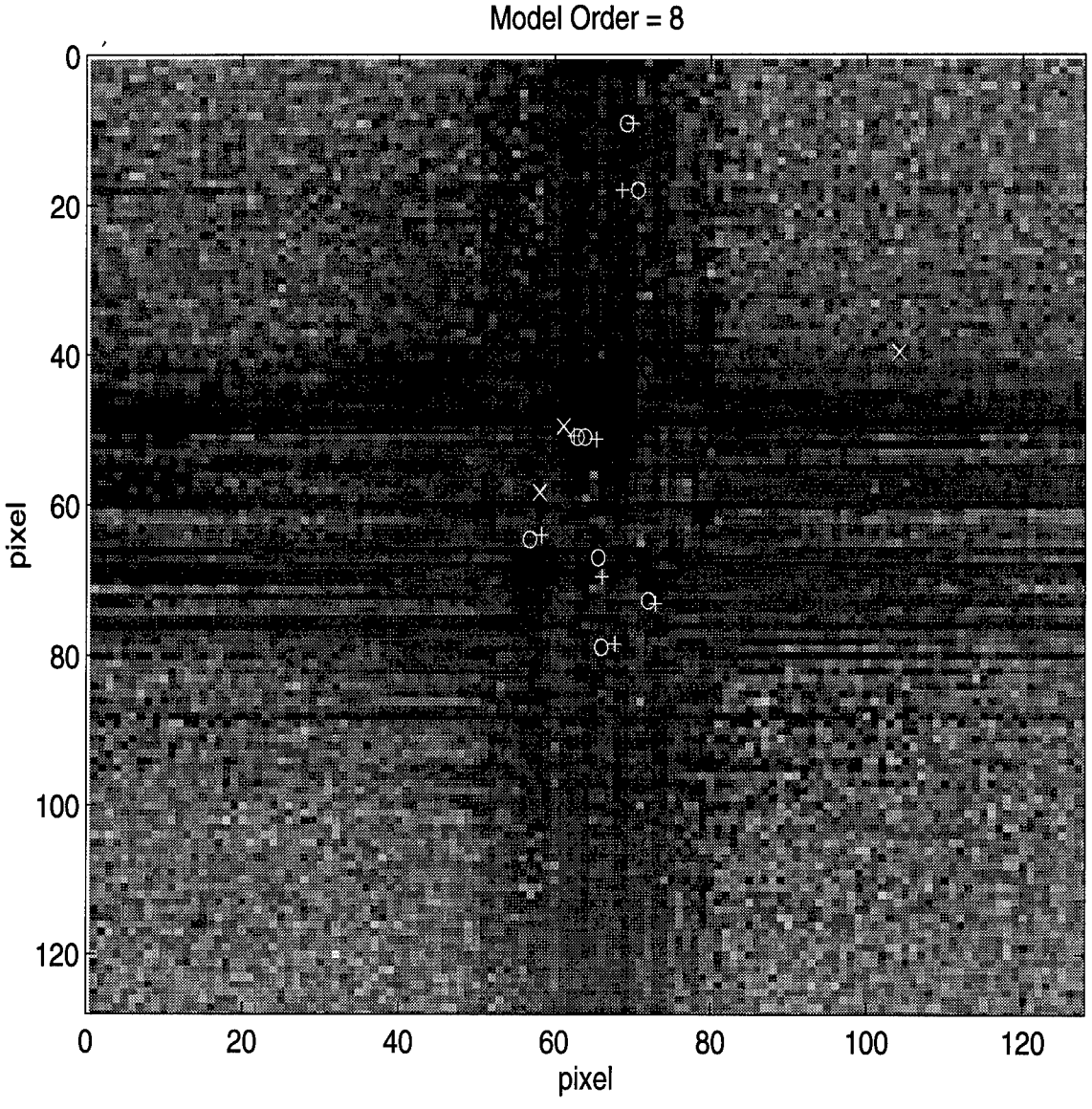


Figure 63. Deterministic ML scatterers (16 × 16 C-29 inverse distance focused image).
o = 2D IQML + = 2D UCIQML x = 2D ARIQML

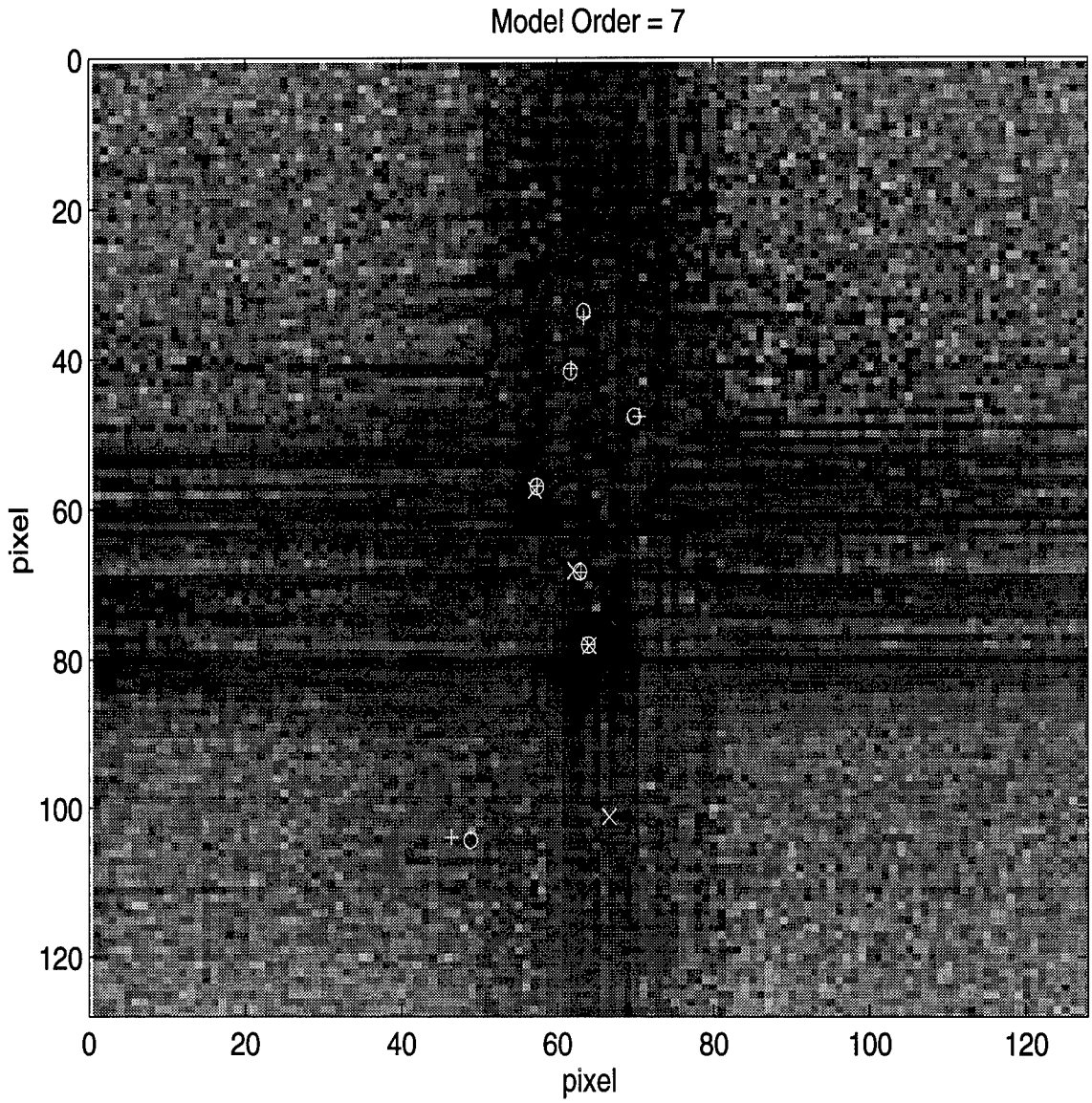


Figure 64. Deterministic ML scatterers ($16 \times 16$ C-29 scattering center focus image).
o = 2D IQML + = 2D UCIQML x = 2D ARIQML

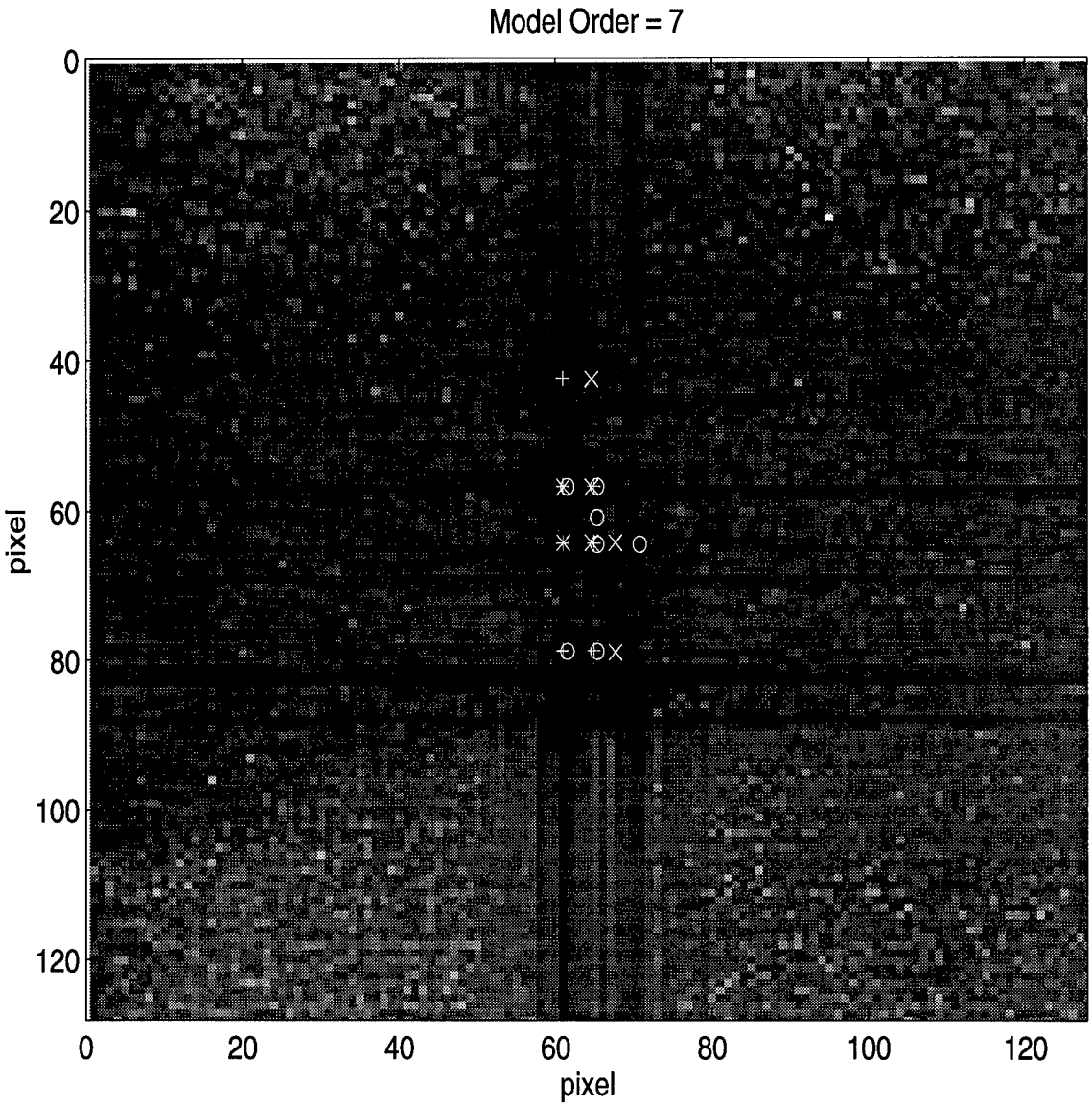


Figure 65. Stochastic ML scatterers ($16 \times 16$ tank image).
o = 2D MODE + = 2D UCMODE x = 2D ARMODE

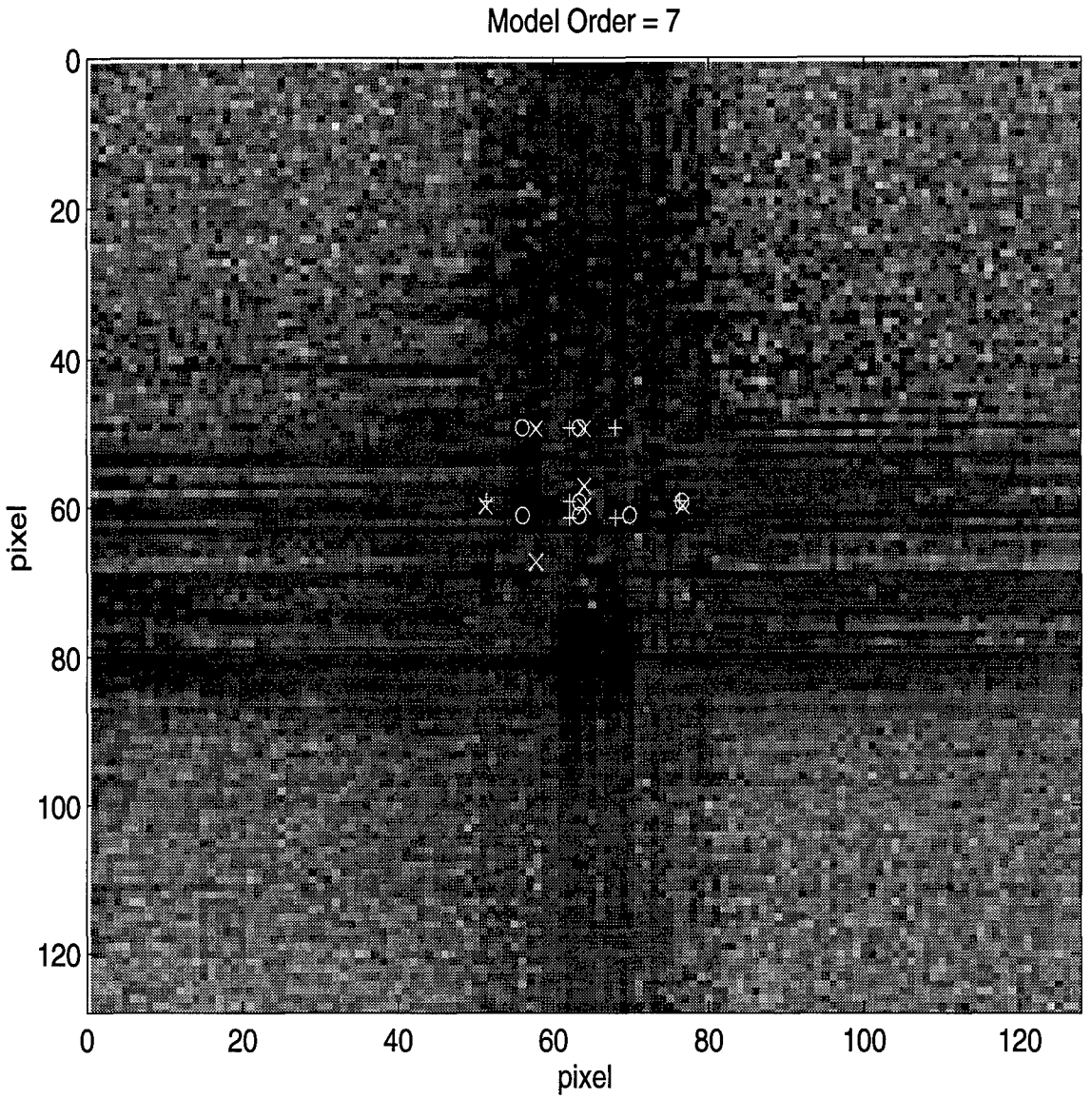


Figure 66. Stochastic ML scatterers ($16 \times 16$ C-29 inverse distance focused image).
o = 2D MODE + = 2D UCMODE x = 2D ARMODE

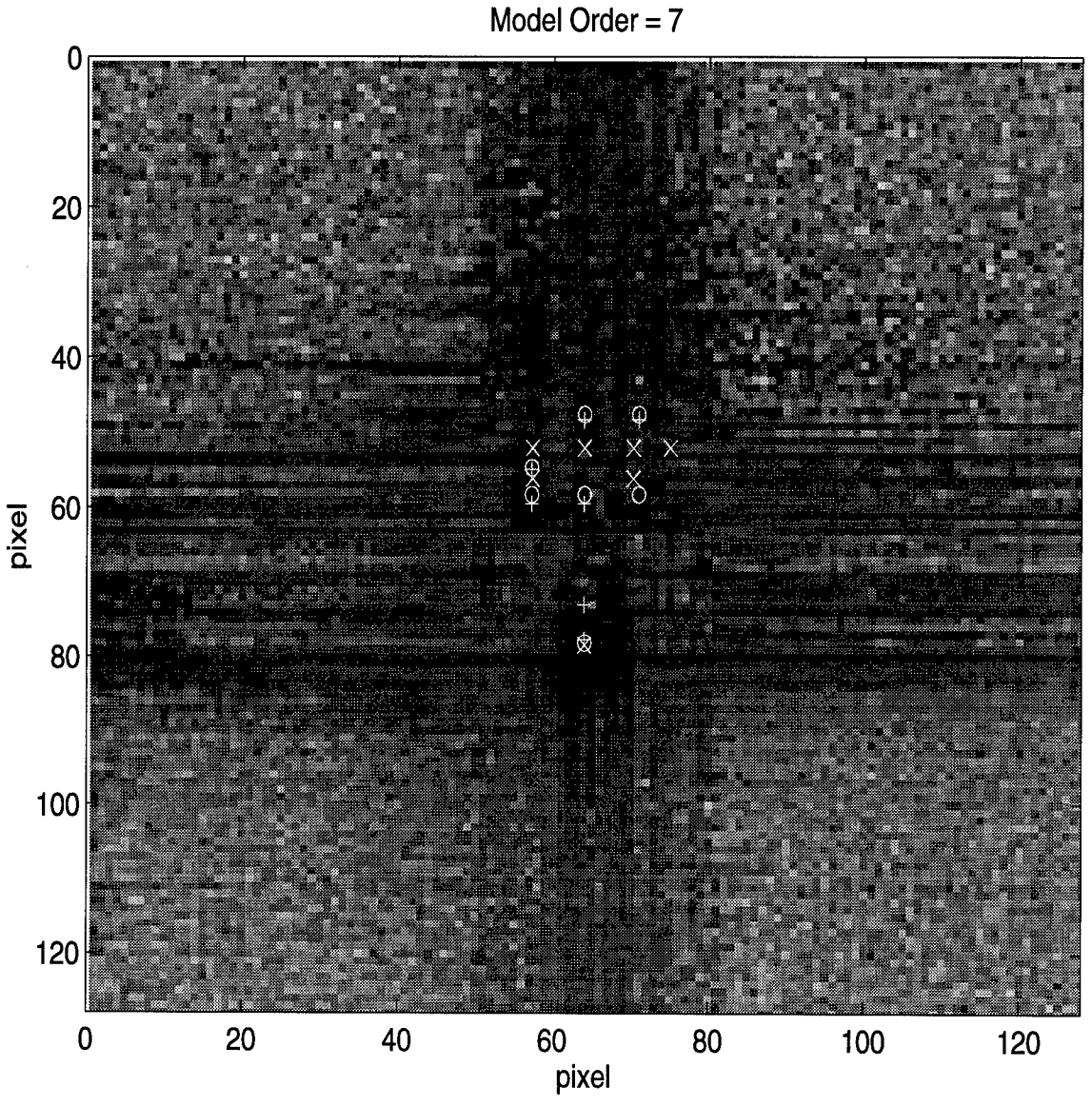


Figure 67. Stochastic ML scatterers (16 × 16 C-29 scattering center focused image).
o = 2D MODE + = 2D UCMODE x = 2D ARMODE

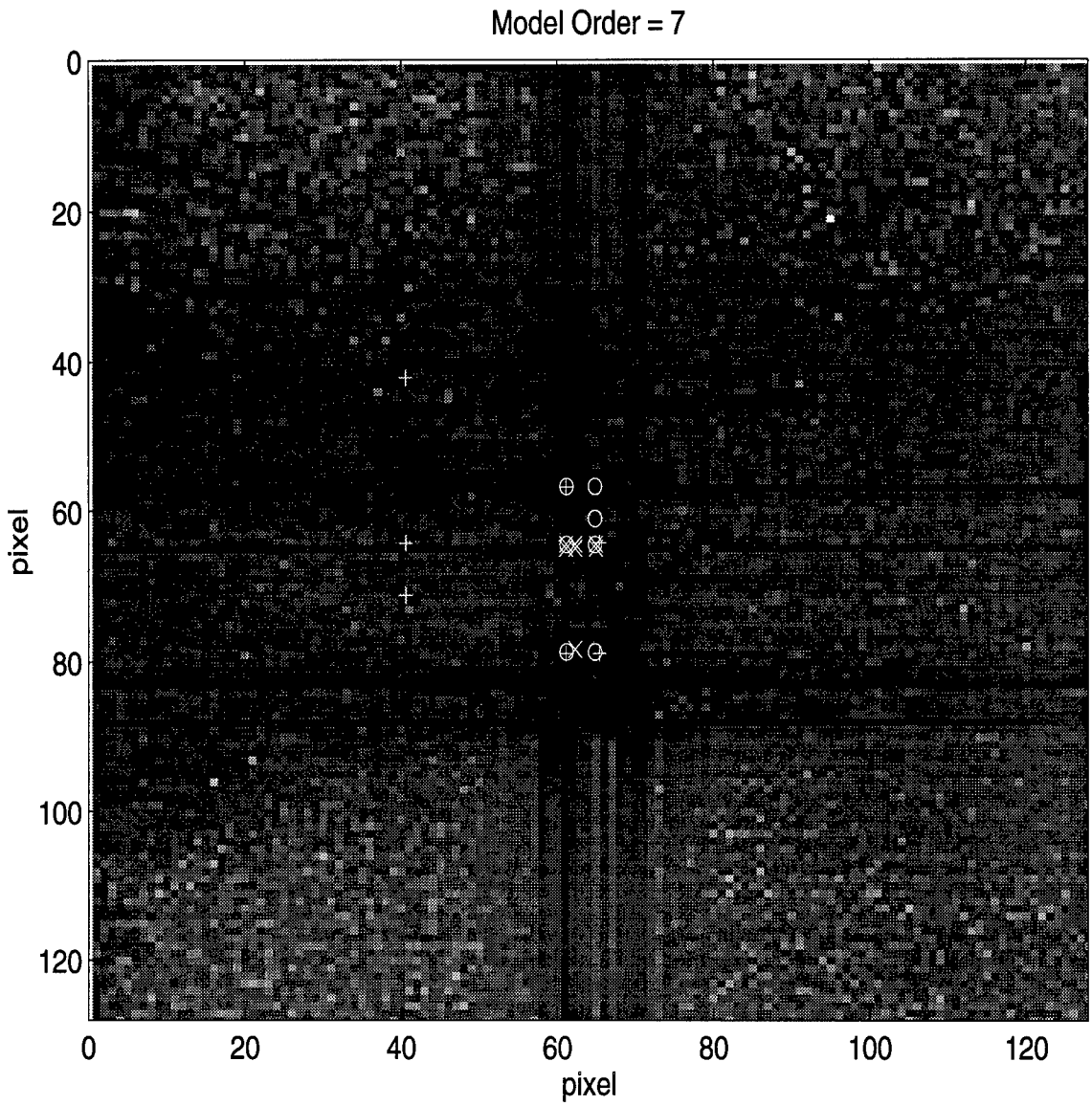


Figure 68. Scattering centers of other 2-D techniques (16 × 16 tank image).
o = 2D PRONY + = STATE SPACE x = MEMP

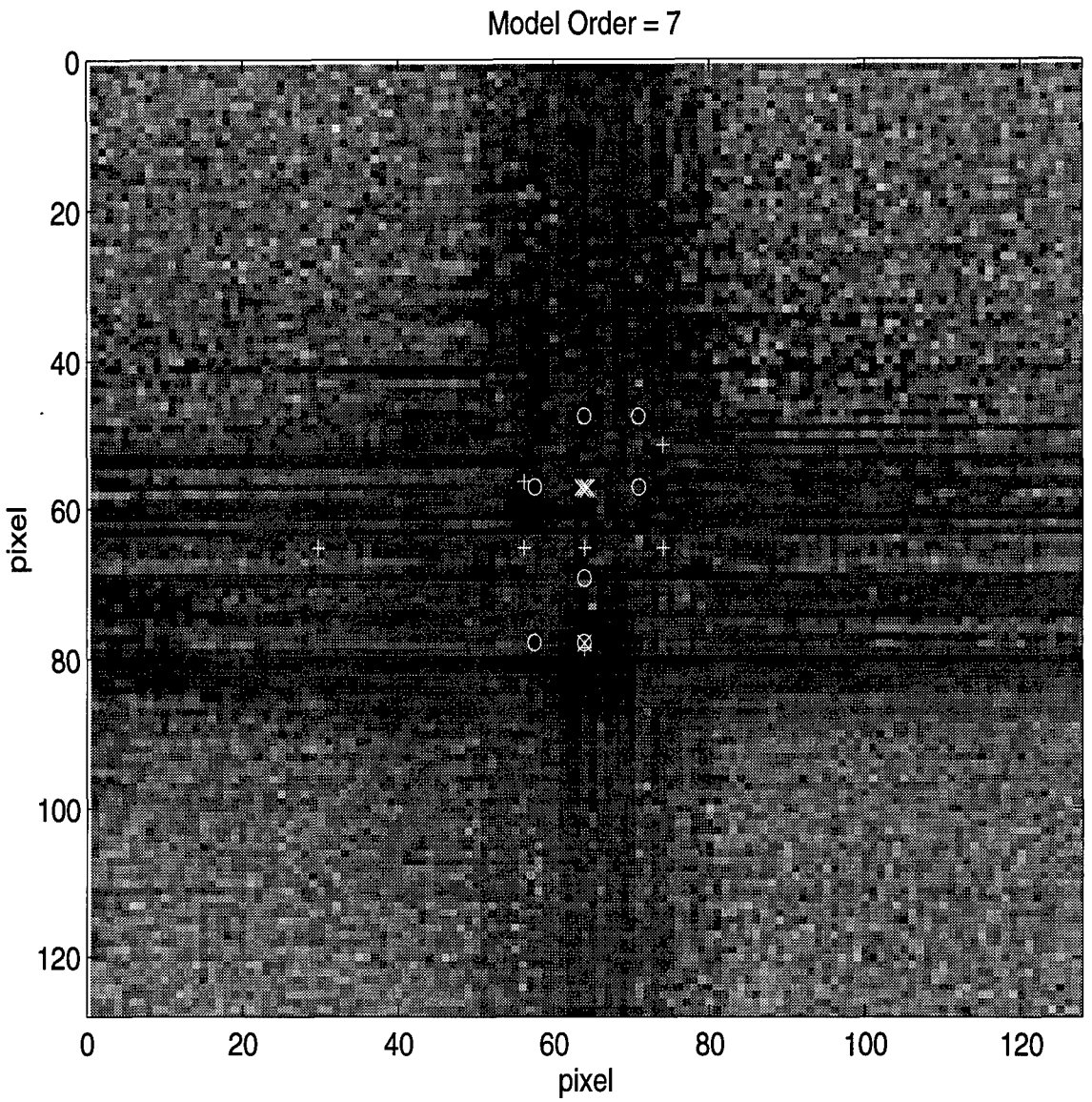


Figure 69. Scattering centers of other 2-D techniques (16 × 16 C-29 scattering center focused image).
o = 2D PRONY + = STATE SPACE x = MEMP

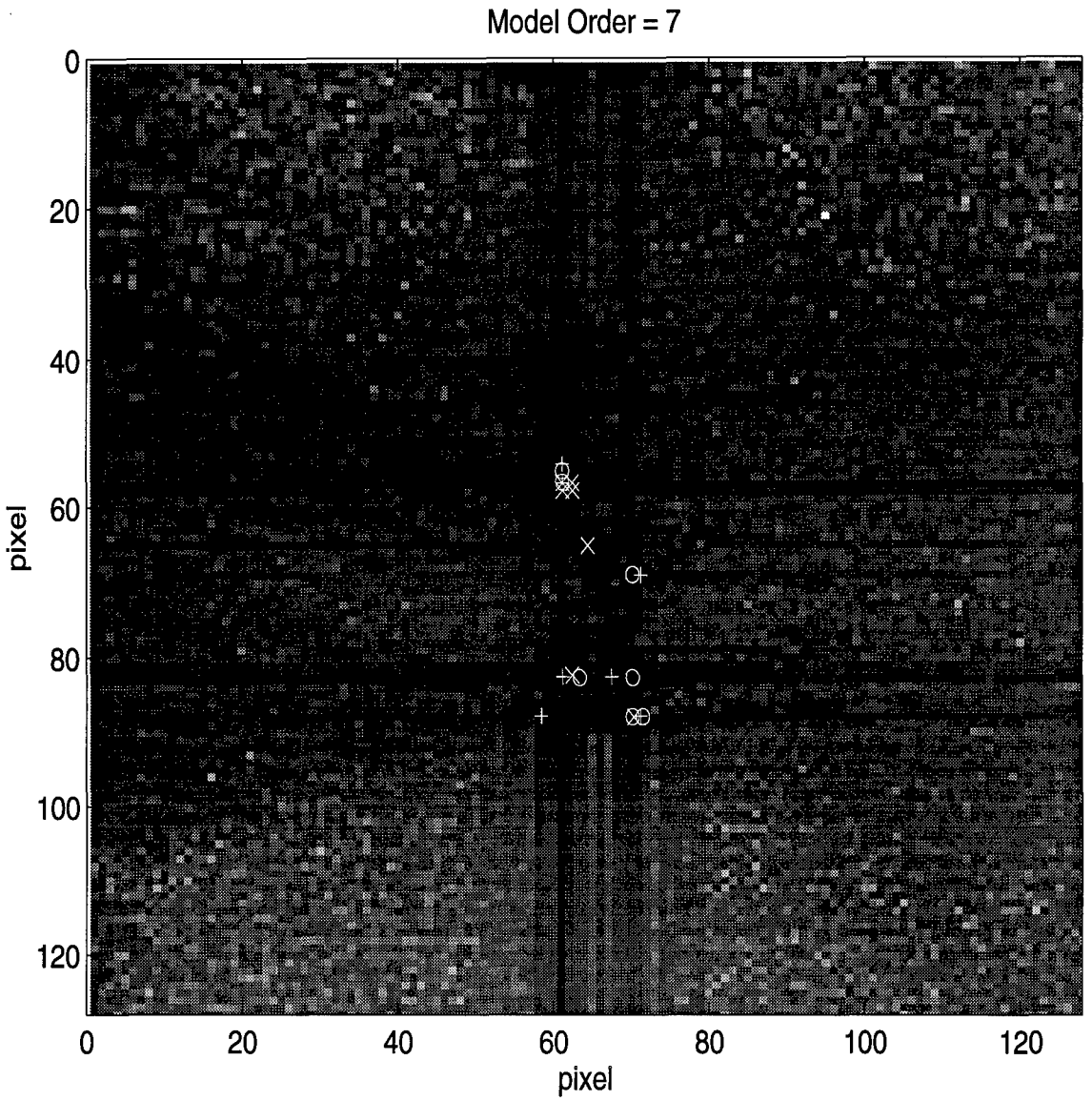


Figure 70. Stochastic ML scatterers (128 × 128 tank image).
o = 2D MODE + = 2D UCMODE x = 2D ARMODE

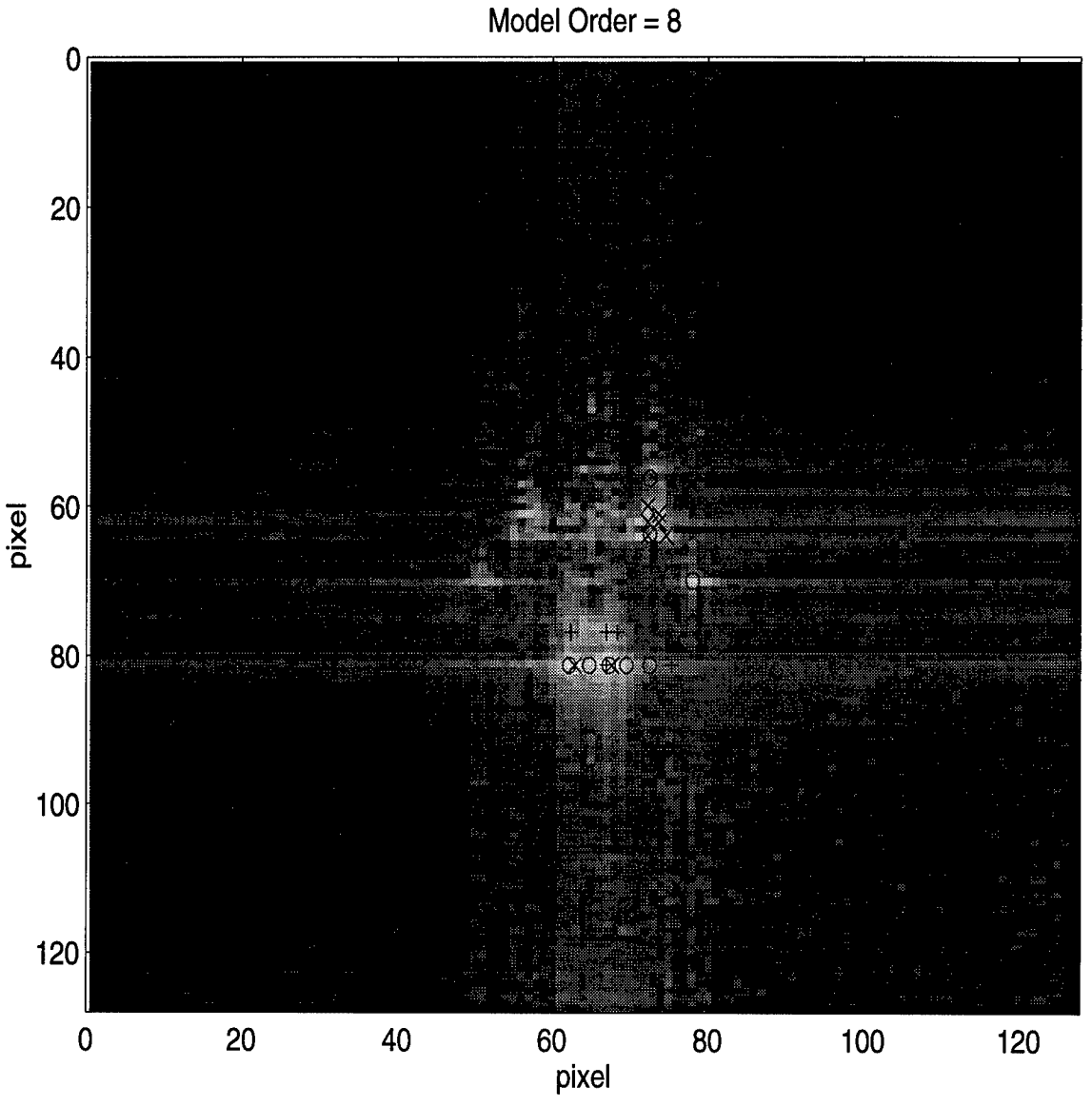


Figure 71. Stochstic ML scatterers (128 × 128 C-29 unfocused image).
o = 2D MODE + = 2D UCMODE x = 2D ARMODE

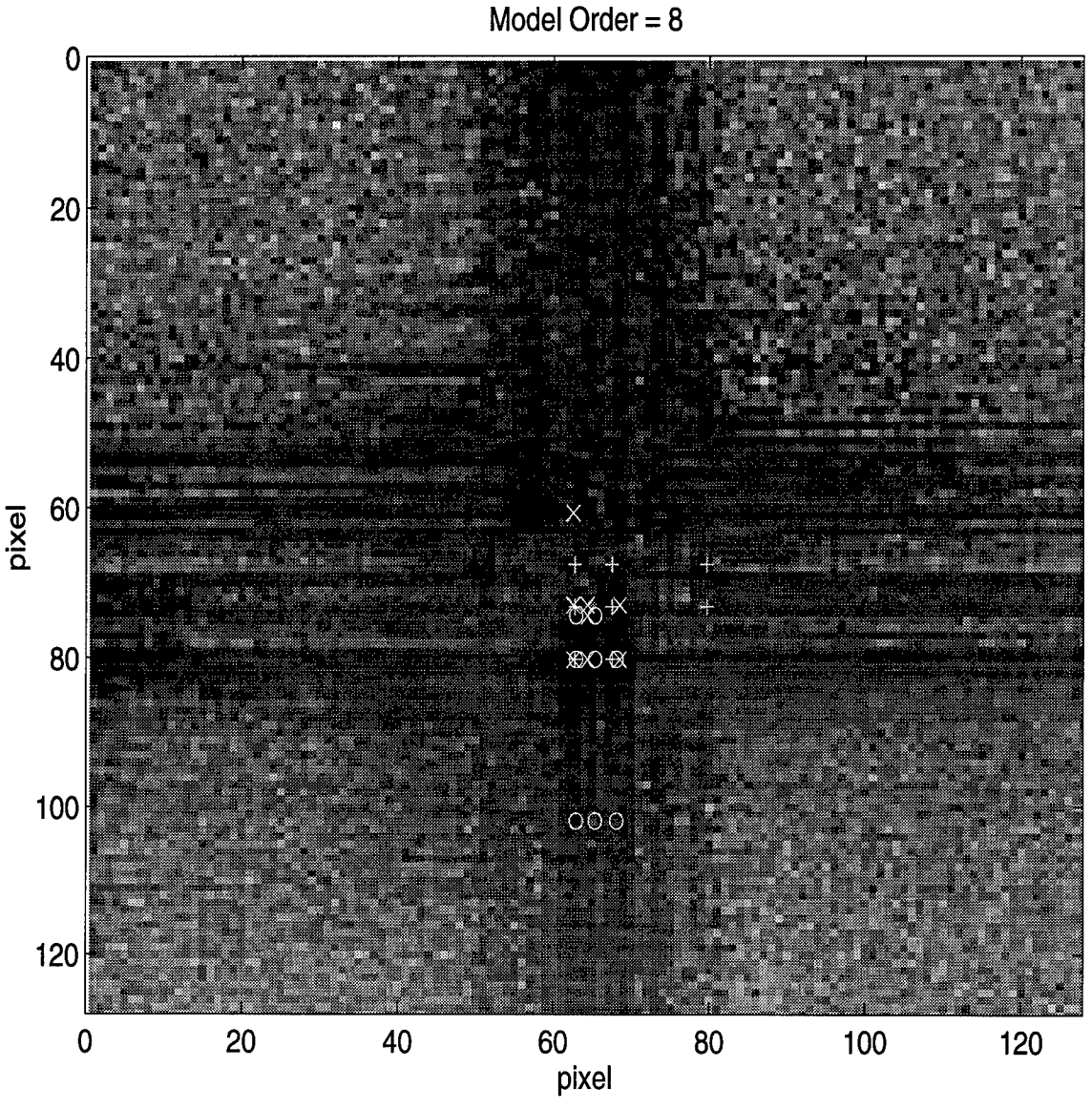


Figure 72. Stochastic ML scatterers (128 × 128 C-29 scattering center focused image).
o = 2D MODE + = 2D UCMODE x = 2D ARMODE

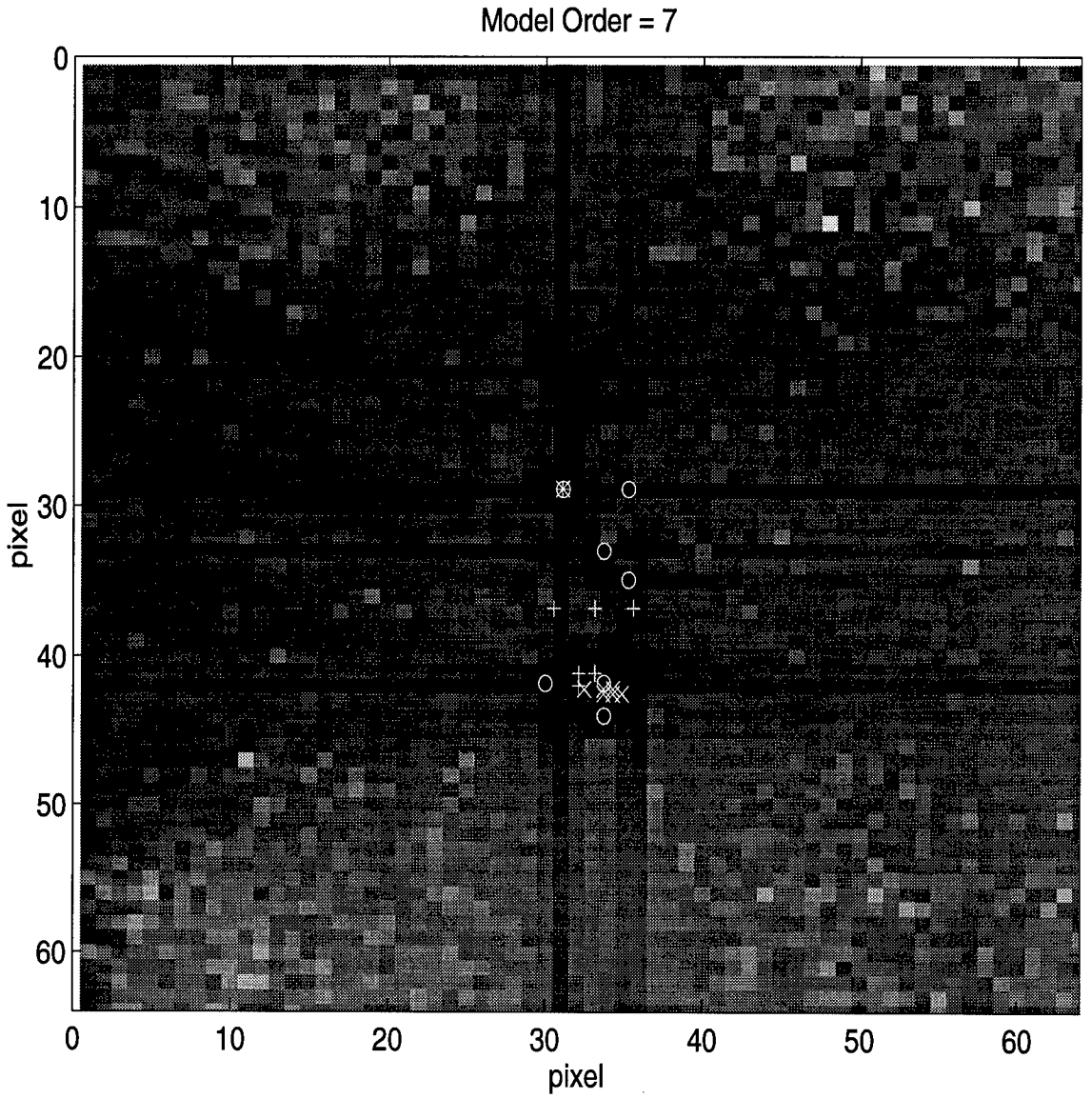


Figure 73. Scattering centers of other 2-D techniques (64 × 64 tank image).
o = 2D PRONY + = STATE SPACE x = MEMP

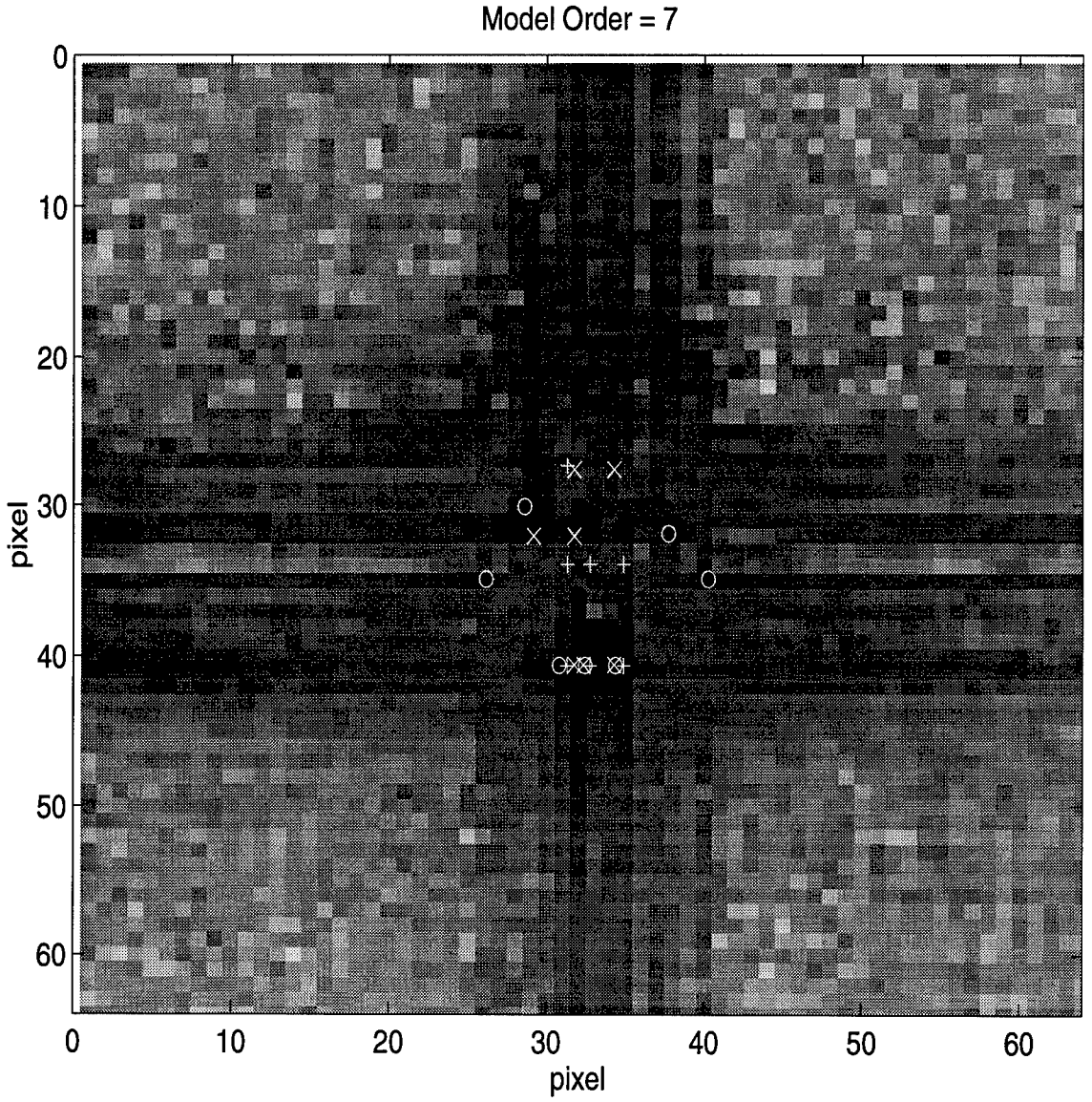


Figure 74. Scattering centers of other 2-D techniques (64 × 64 C-29 scattering center focused image).
o = 2D PRONY + = STATE SPACE x = MEMP

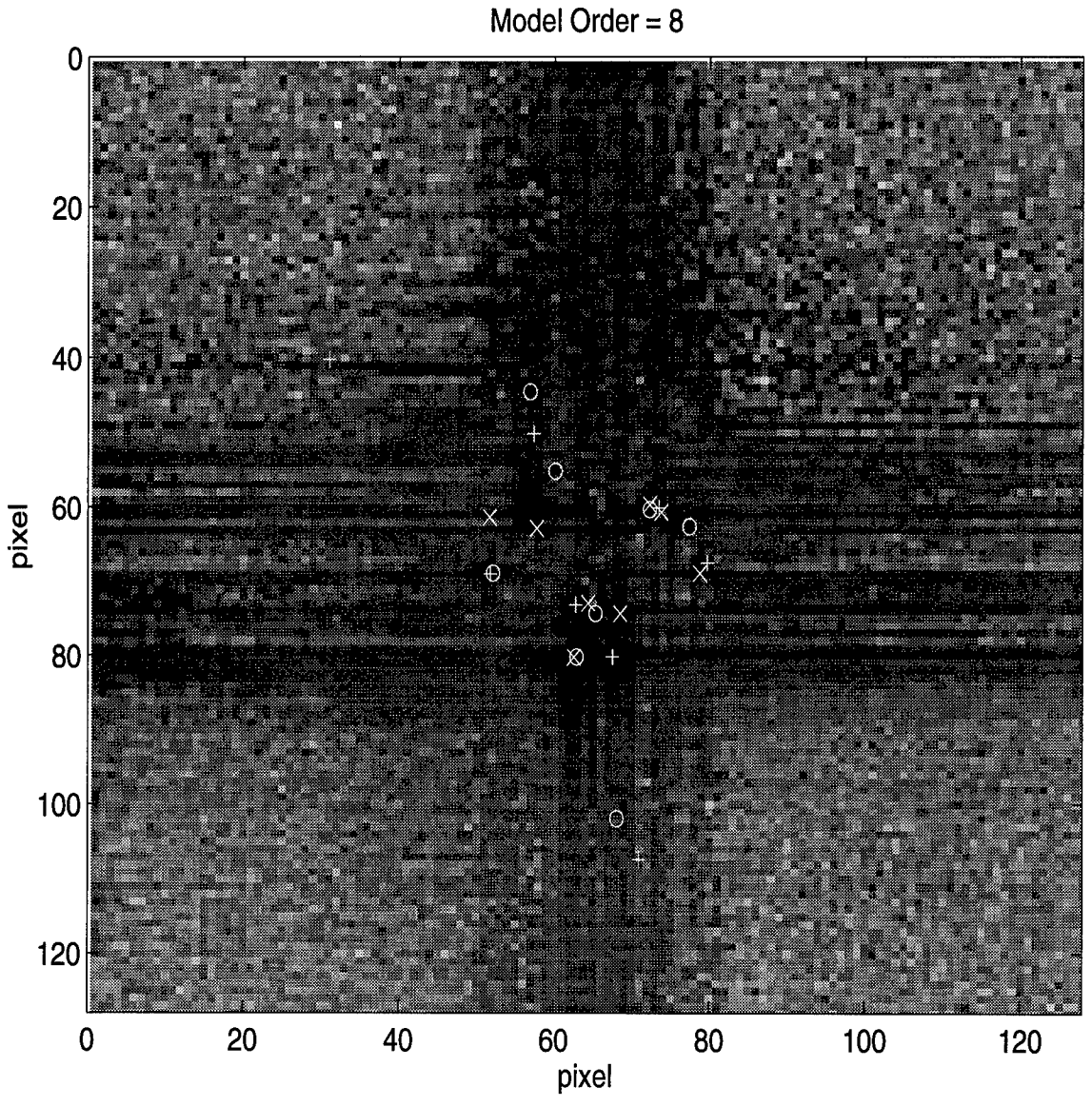


Figure 75. Stochastic ML distinct scatterers (128 × 128 C-29 scattering center focused image). o = 2D MODE + = 2D UCMODE x = 2D ARMODE

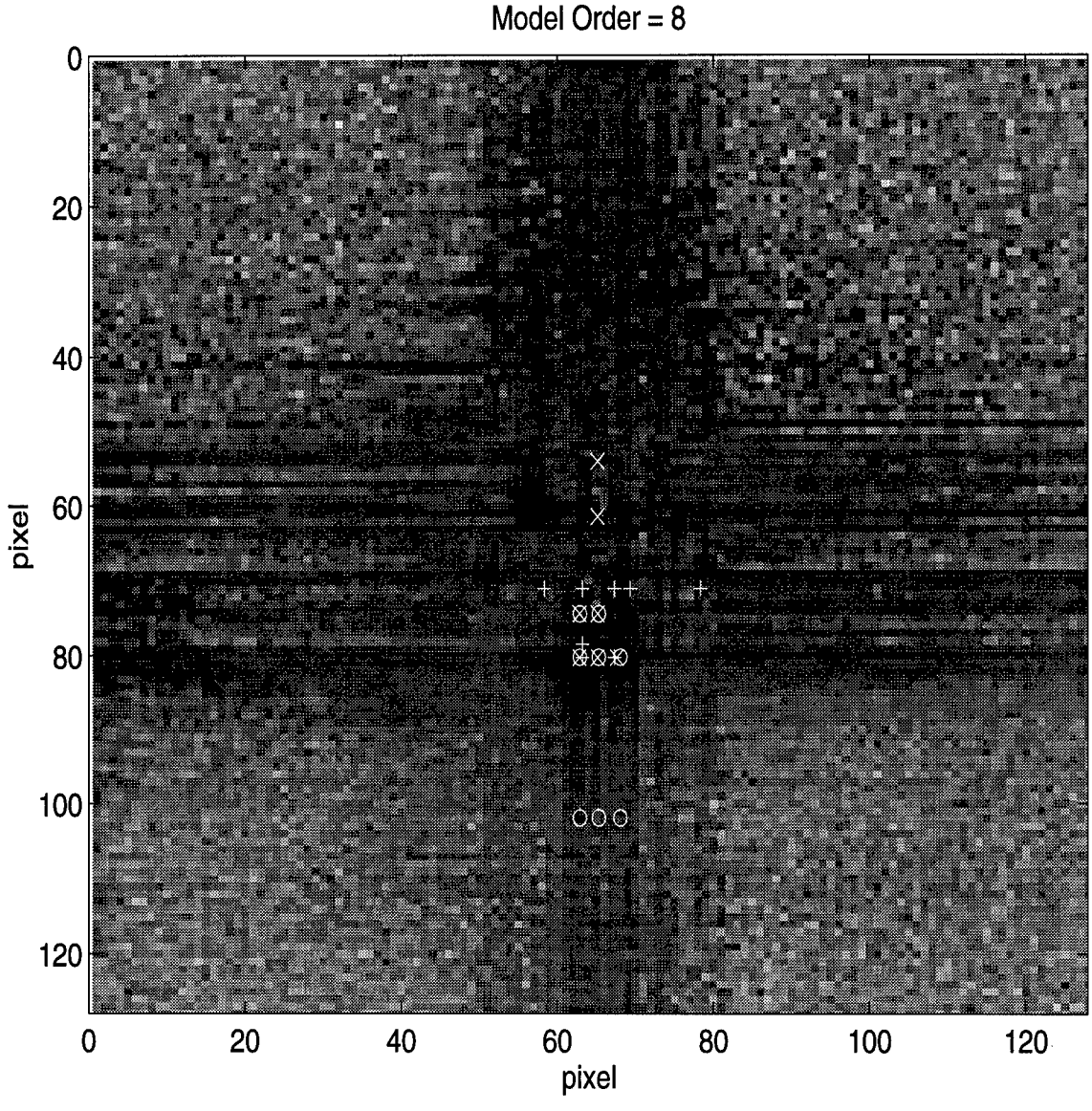


Figure 76. Stochastic ML non-distinct scatterers ($128 \times 128$ C-29 scattering center focused image). o = 2D MODE + = 2D UCMODE x = 2D ARMODE

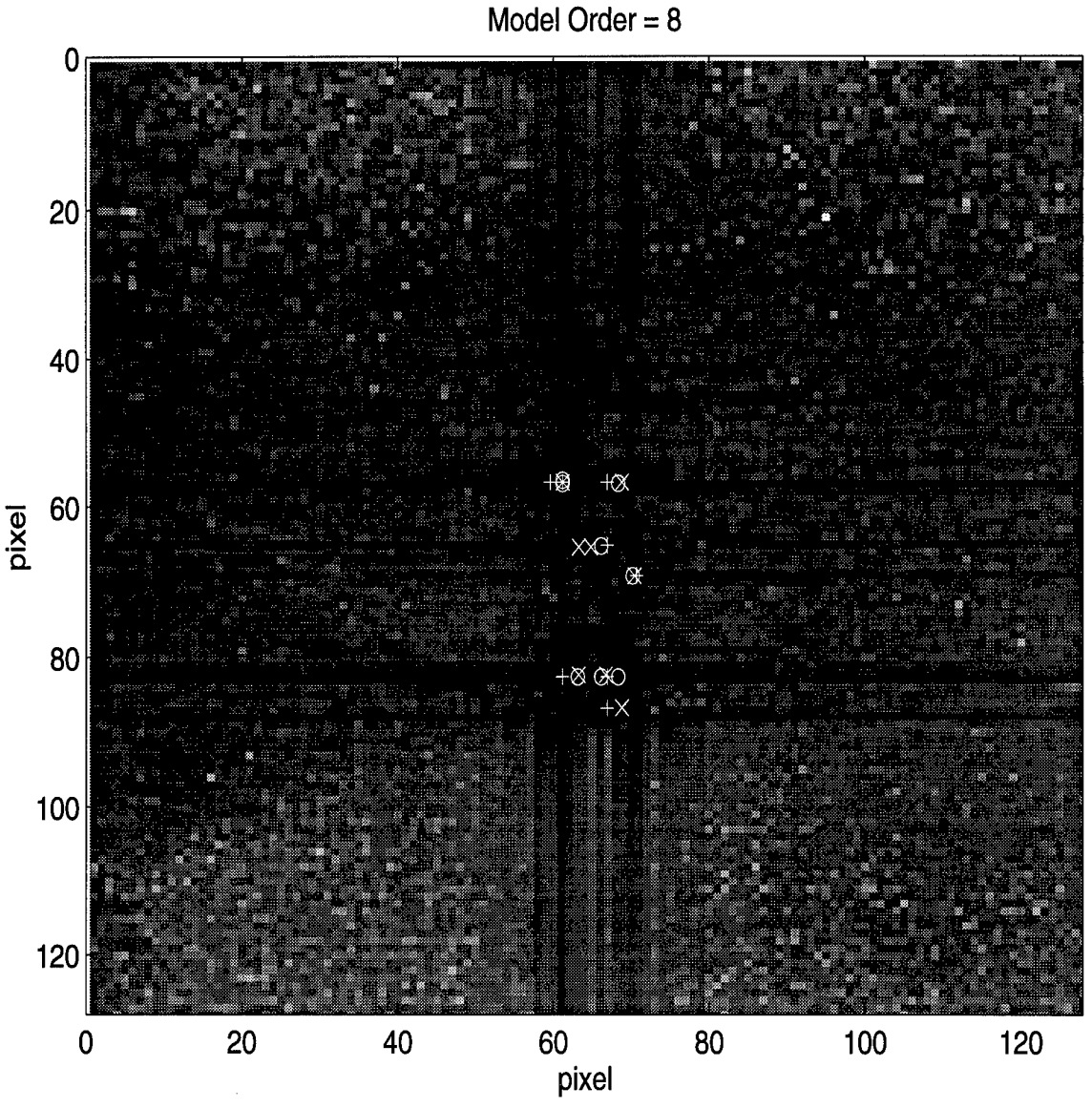


Figure 77. Stochastic ML scatterers, 2D Prony matching (64 × 64 Tank image).
o = 2D MODE + = 2D UCMODE x = 2D ARMODE

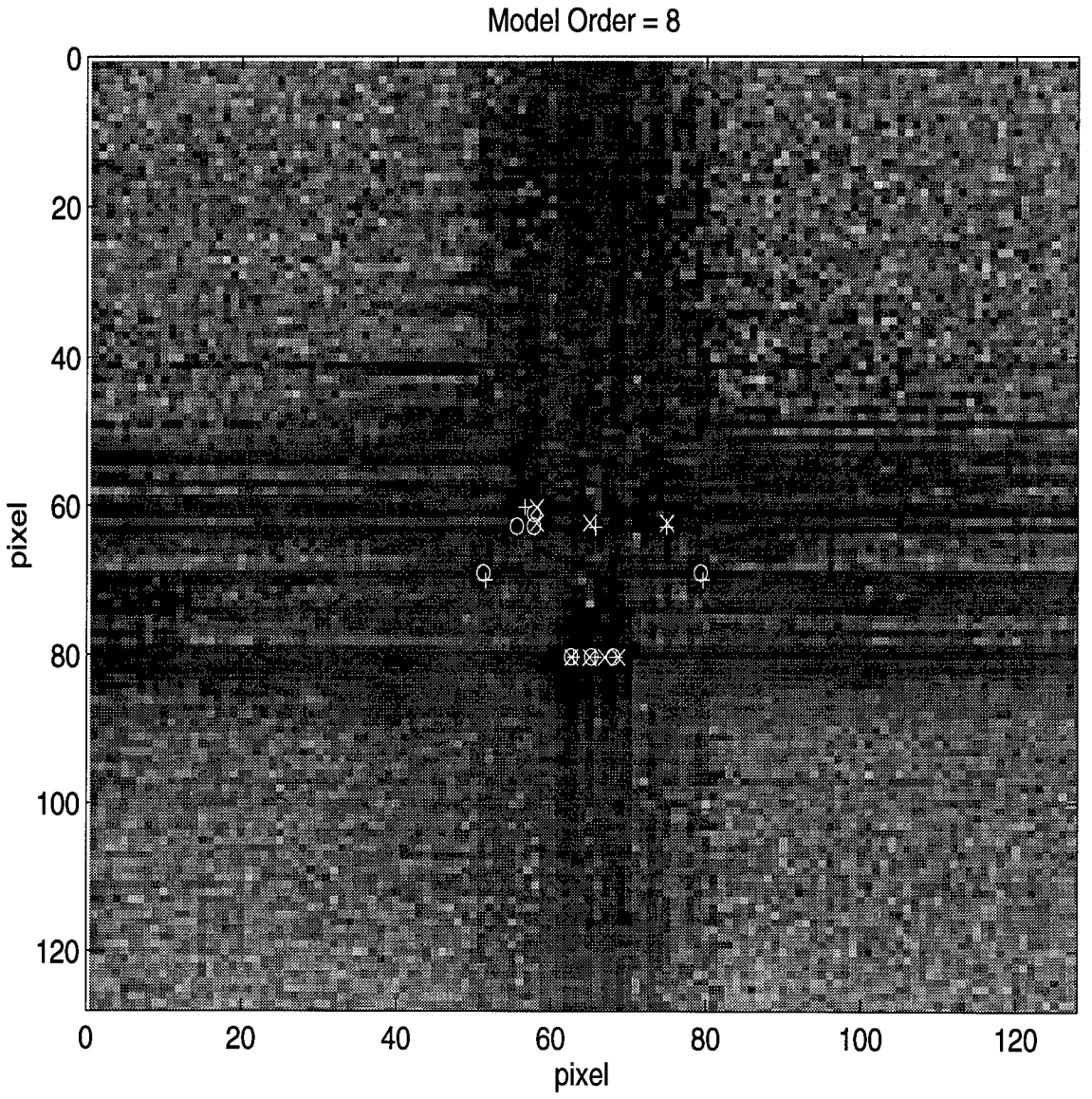


Figure 78. Stochastic ML scatterers, 2D Prony matching ($64 \times 64$ C-29 scattering center focused image).
o = 2D MODE + = 2D UCMODE x = 2D ARMODE

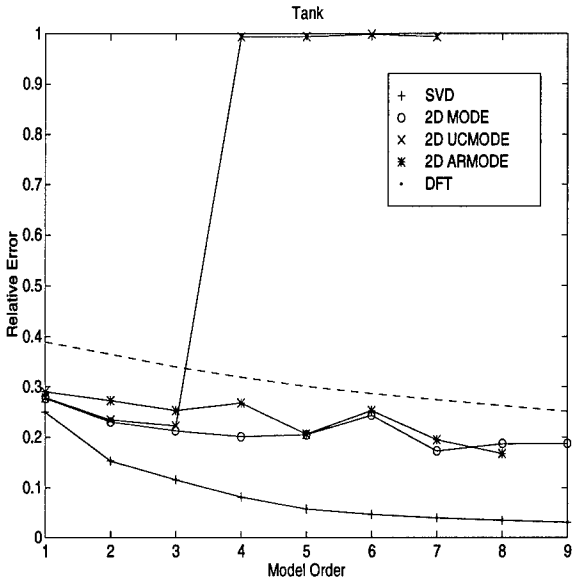


Figure 79. Representation error vs. model order (Stochastic ML, $64 \times 64$ tank image), 2D Prony matching. Techniques: 2-D method of direction estimation (2D MODE), unknown colored noise mode2d (2D UCMODE), and exponential noise ucmode2d (2D ARMODE), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

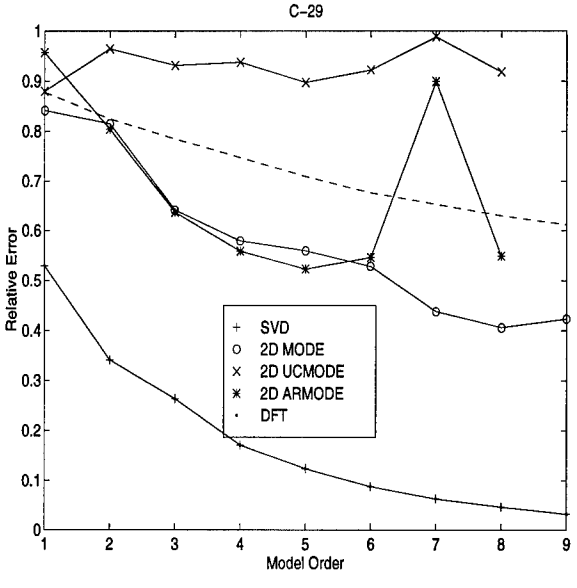


Figure 80. Representation error vs. model order (Stochastic ML, $64 \times 64$ C-29 scattering center focused image), 2D Prony matching. Techniques: 2-D method of direction estimation (2D MODE), unknown colored noise mode2d (2D UCMODE), and exponential noise ucmode2d (2D ARMODE), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

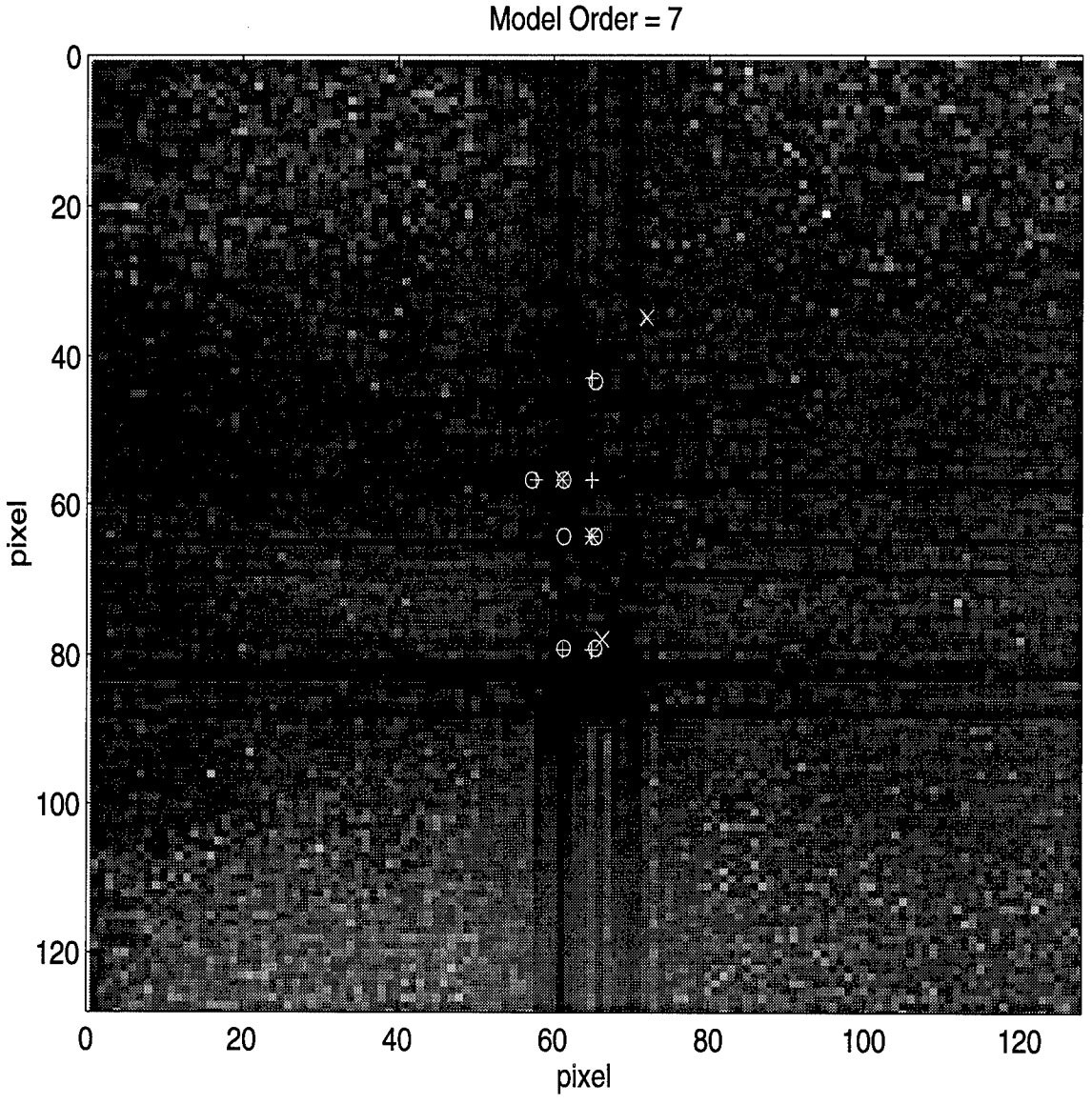


Figure 81. Deterministic ML scatterers, 2D Prony matching (16 × 16 Tank image).
o = 2D IQML + = 2D UCIQML x = 2D ARIQML

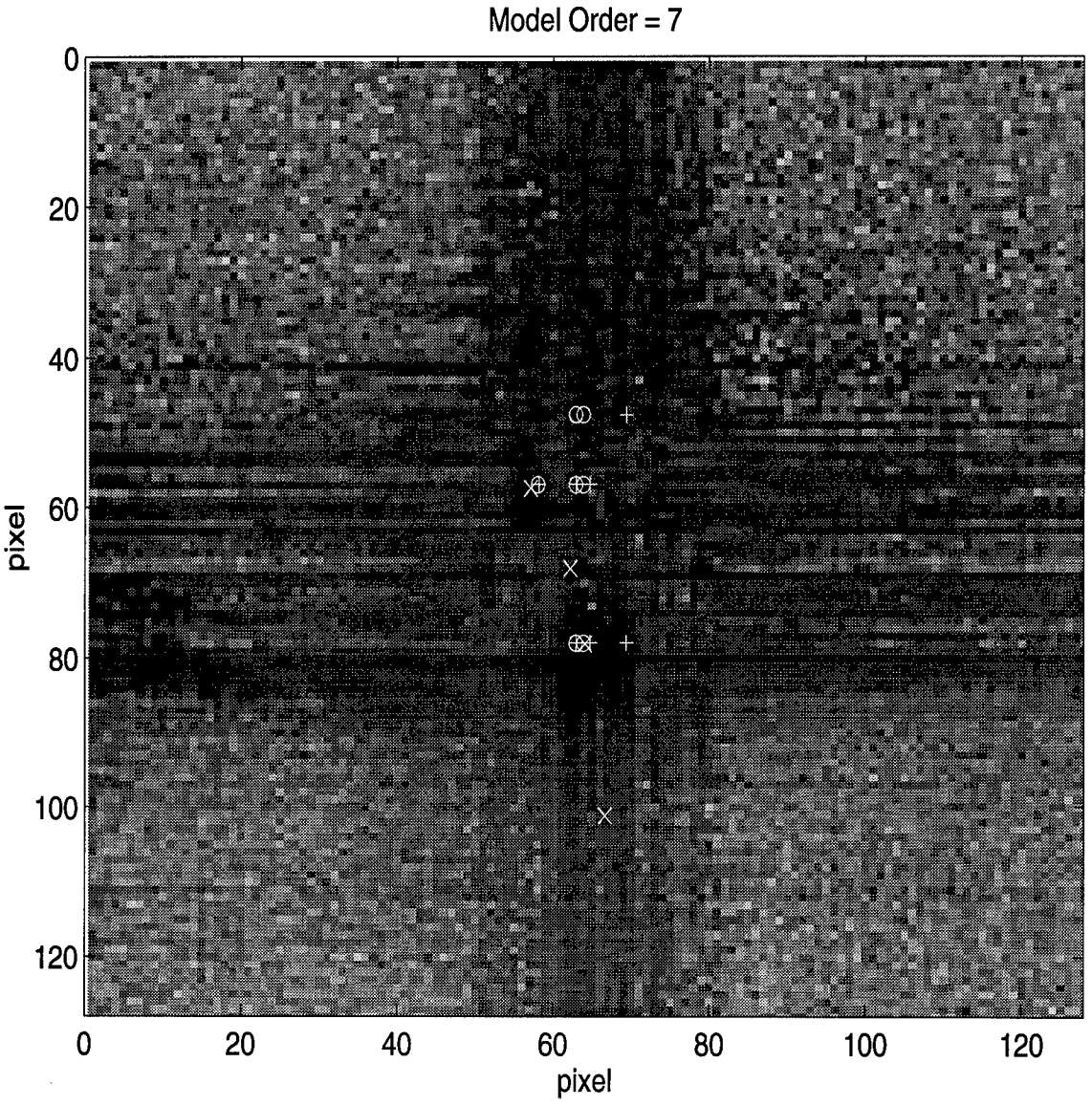


Figure 82. Deterministic ML scatterers, 2D Prony matching ($16 \times 16$ C-29 scattering center focused image).
o = 2D IQML + = 2D UCIQML x = 2D ARIQML

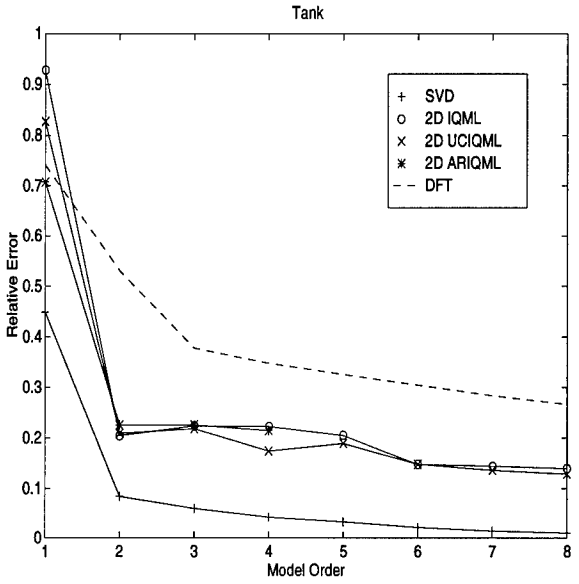


Figure 83. Representation error vs. model order (Deterministic ML, 16 × 16 tank image), 2D Prony matching. Techniques: 2-D iterative quadratic maximum likelihood (2D IQML), unknown colored noise iqml2d (2D UCIQML), and exponential noise uciqml2d (2D ARIQML), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

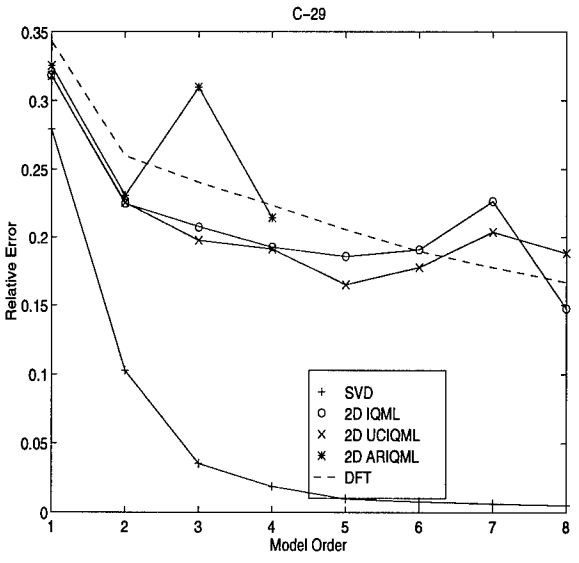


Figure 84. Representation error vs. model order (Deterministic ML, 16×16 C-29 scattering center focused image), 2D Prony matching. Techniques: 2-D iterative quadratic maximum likelihood (2D IQML), unknown colored noise iqml2d (2D UCIQML), and exponential noise uciqml2d (2D ARIQML), Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT)

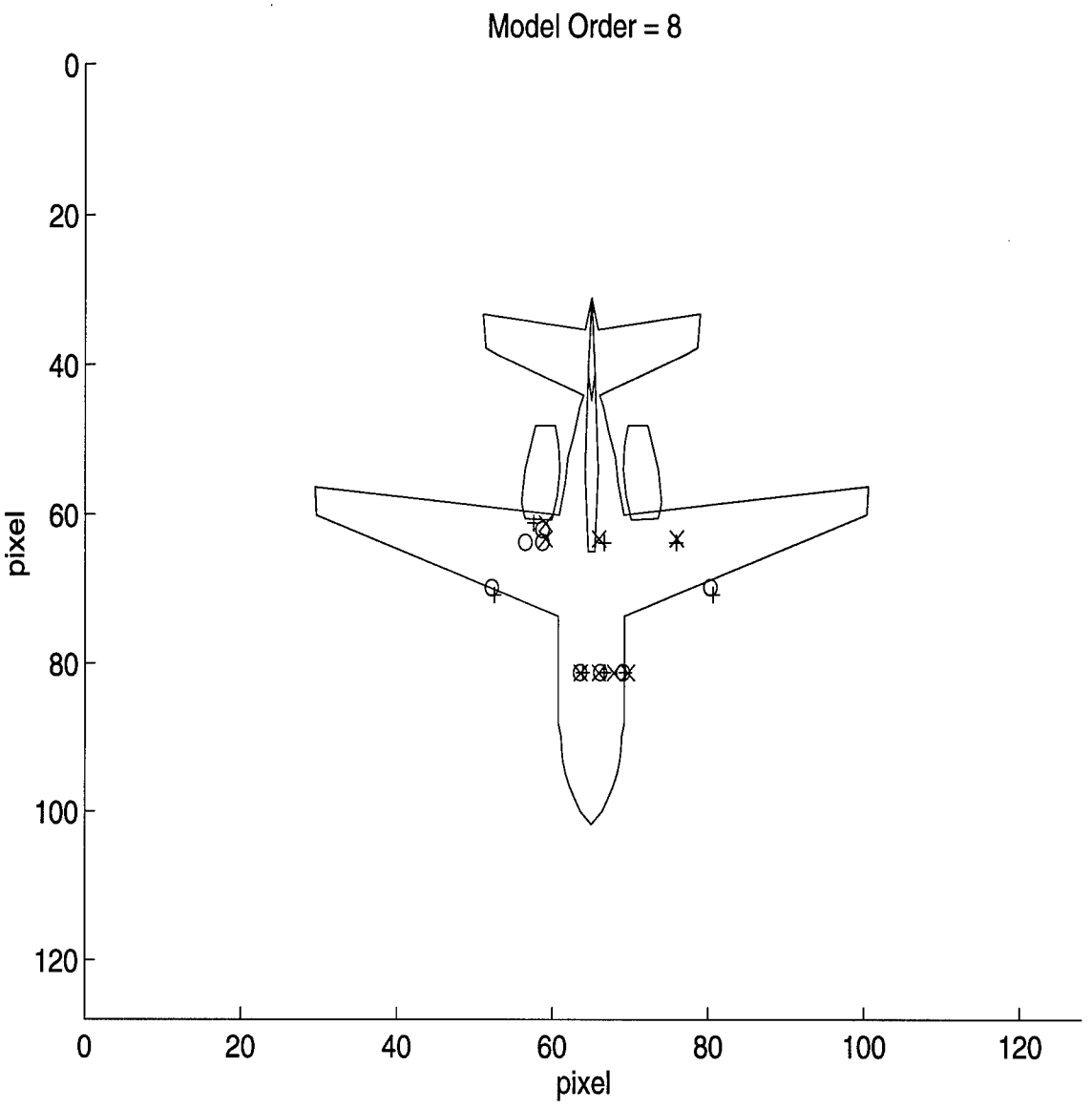


Figure 85. Stochastic ML scatterers, 2D Prony matching (64 × 64 C-29 scattering center focused image).
o = 2D MODE + = 2D UCMODE x = 2D ARMODE

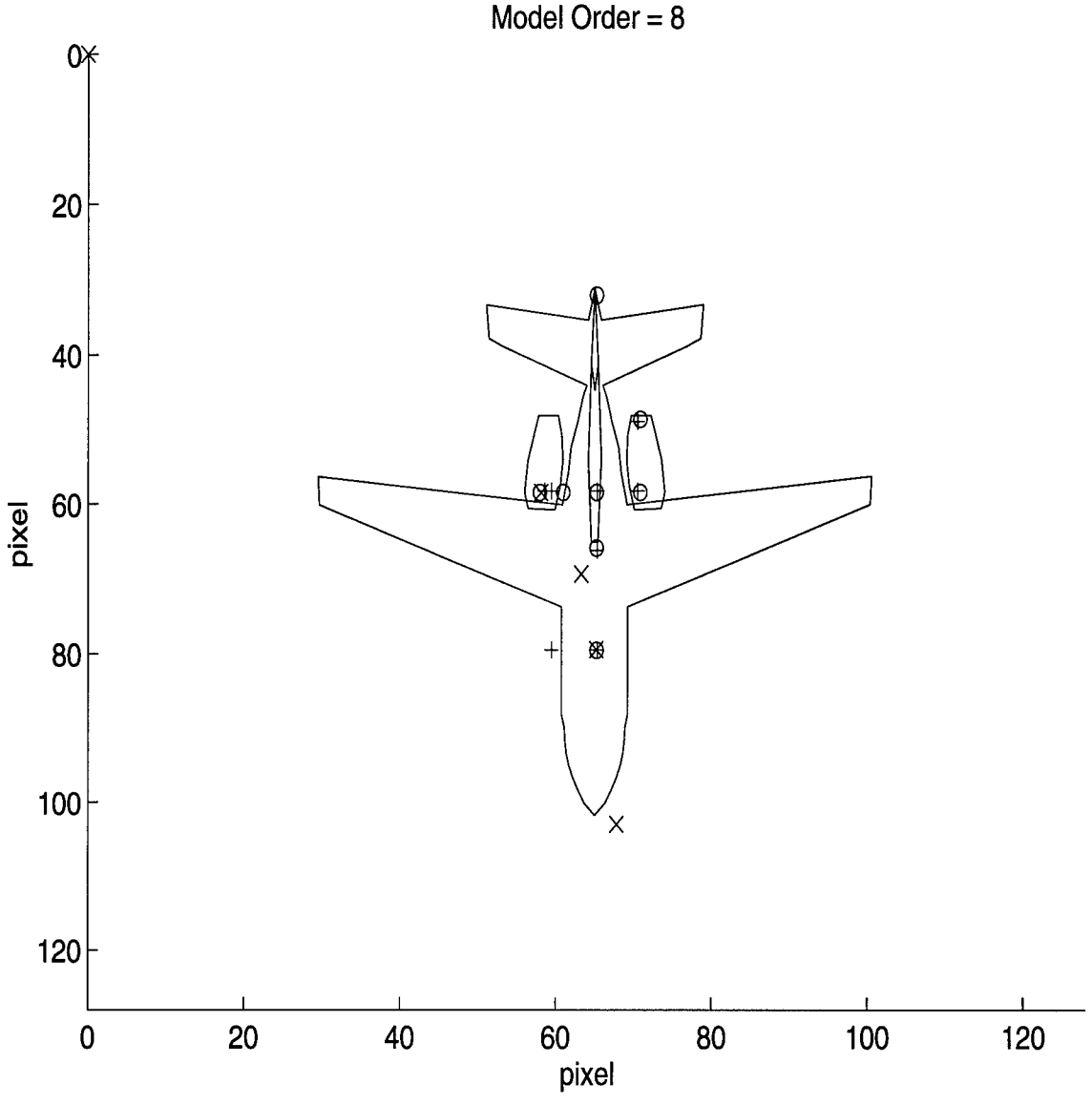


Figure 86. Deterministic ML scatterers, 2D Prony matching (16 × 16 C-29 scattering center focused image).
o = 2D IQML + = 2D UCIQML x = 2D ARIQML

# *VIII. Conclusions*

Maximum likelihood techniques were developed to estimate exponentials in colored noise. Methods were developed for one and two dimensions, for stationary colored noise, and colored noise modeled with a rational polynomial. For all these cases, separate methods were developed for a deterministic signal assumption and for a stochastic signal assumption.

All these techniques perform well when the noise model matches the actual noise. In these cases, accuracy in estimating exponential frequencies was improved over techniques assuming white noise. All techniques attained the Cramér-Rao estimation bound (CRB) either for short data records or asymptotically with good performance for short data records. These techniques also performed well when the noise was white and, in some cases, improved low SNR performance by mitigating the colored appearance of a finite sample of white noise. In two dimensions, the increased degrees of freedom of 2-D data makes it difficult to accurately estimate the noise. Thus, only slight improvement over the white noise techniques was observed and no technique attained the CRB for short data records.

The colored noise techniques were applied to Synthetic Aperture Radar data which, when correctly processed, can be modeled as the sum of damped exponentials for targets consisting of point scatterers. In one dimension, colored noise techniques corrected for a poor choice of model order in much the same way as modeling the data with an increased model order and discarding the low energy part (overmodeling). In fact, the 1-D colored noise estimation techniques can be viewed as overmodeling then using the overmodeled part as a noise estimate and to improve the estimate of the signal. In two dimensions, better fit of the data at low model orders was not observed. The 2-D techniques based on 2D IQML identified a limited number of scatterer in one dimension due to a limitation on the algorithm to distinct frequencies in one dimension. The 1-D by 1-D techniques based on 2D MODE tended to identify scatterers along lines in either dimension. The use of 2-D Prony matching improved scattering center locations in all techniques. The 2-D techniques based on 2D IQML and on 2D MODE when paired with the matching techniques of 2D Prony consistently identified many individual scatterers and had few false scatterers. In these cases, the unknown colored noise techniques identified different scatterers from the white noise techniques and fit the data better as well.

The 2-D techniques based on the colored noise model more accurately model SAR data than existing 2-D white noise techniques. The 2-D colored noise techniques developed in this thesis should significantly aid in the development of model-based pattern recognition algorithms to identify targets in SAR images. All of the 2-D techniques when combined with scattering center focus and 2D Prony matching provide robust tools for identifying scattering centers on an object in a SAR image. These techniques will aid in understanding the consistency with which scatterers can be identified across different SAR images. Of the techniques the 2D IQML and 2D UCIQML techniques tied with 2D Prony matching appear to give the best estimates of the locations of true 2-D scatterers while avoiding overmodeling of these scatterers.

## *8.1 Contributions*

List of the contributions in this dissertation.

- A concise mathematical model for damped exponentials on an irregularly sampled grid. (Chapter 4)
- A computationally efficient (real-time) method for interpolating or focusing SAR images containing damped exponentials. (Chapter 4)
- Closed-form solution of the 1-D maximum likelihood problem of estimating deterministic exponentials in unknown colored noise. (Chapter 5)
- Reformulation of the 1-D colored noise stochastic maximum likelihood problem to avoid a global search for the solution. (Chapter 5)
- Several 1-D deterministic and stochastic maximum likelihood algorithms for estimating 1-D exponentials in colored noise. (Chapter 5)
- Extension of deterministic maximum likelihood estimation of 2-D exponentials to unknown colored noise. (Chapter 6)
- Extension of stochastic maximum likelihood estimation of 2-D exponentials on a 1-D by 1-D grid to unknown colored noise. (Chapter 6)
- A spectral model to produce unique 2-D spectra (noise, or an impulse driving a 2-D damped exponential filter). (Chapter 6)

- Several 2-D deterministic and stochastic maximum likelihood algorithms for estimating 2-D exponentials in colored noise (Chapter 6)
- Unique properties of lower (upper) triangular Toeplitz matrices, commutivity and a lower (upper) triangular Toeplitz inverse. (Appendix A)

### *8.2 Future Work*

There are several areas where the work initiated by this dissertation could be expanded and potential new insights gained:

- Apply the unknown colored noise techniques developed in this dissertation to multiple data instances and unit-circle constrained frequencies. In this case, many other application are available for the techniques, such as, array processing in the presence of noisy jammers.
- Develop methods of simultaneously estimating the signal and the noise for the unknown colored noise exponential estimation methodology developed in this dissertation. Such methods will likely lead to efficient techniques with respect to the CRB for exponential and noise estimation.
- Investigate the consistency and resolution with which scatterers are identified at different look angles to the target. Recognition of different targets requires consistent identification of the scatterer locations on the targets. The large set of XPATCH (three SAR bands, several targets, several elevations and all aspects), and C-29 chamber data (many elevations and many aspects at very fine frequency and angle resolution) allows for an extensive analysis.
- Develop a model-based pattern recognition system using relative locations of scatterers on the target. Investigation of real world effects such as occlusion of scattering centers will determine the utility of such a system.

## *Appendix A. Commutivity of Operations*

The approximate commutivity of the operations in Equation 105 is based on the commutivity of the convolution operation. Similarly, the matrix formulation of convolution, multiplication of lower (upper) triangular Toeplitz matrices commutes. The commutivity of block-diagonal triangular Toeplitz matrices is implied in [22]. The following theorems establish the commutivity of the triangular Toeplitz matrices, the triangular Toeplitz structure of the inverse of a triangular Toeplitz matrix, and the commutivity of the operations in Equation 105.

**Theorem.** Let $A$ , $B$ be $N \times N$ upper (lower) triangular Toeplitz matrices and their product be $C = AB$, then, the matrices $A$ and $B$ commute, and $C = BA$. *Proof.* The elements of the product of the upper triangular matrices $A$ and $B$ are

$$c_{ij} = \sum_{k=i}^{j} a_{ik} b_{kj} \qquad i \leq j. \tag{214}$$

Since the matrices $A$ and $B$ are also Toeplitz, they are symmetric about the anti-diagonal and the their elements are related as

$$a_{i,k=\{i \ldots j\}} = a_{k=\{j \ldots i\},j}. \tag{215}$$

Thus,

$$c_{ij} = \sum_{k=i}^{j} a_{kj} b_{jk} \qquad i \leq j, \tag{216}$$

or $C = BA$. The lower triangular case is simply seen by taking the transpose of the upper triangular matrix products $AB = BA$.

**Theorem.** Let $T$ be an upper (lower) triangular Toeplitz matrix then its inverse $T^{-1}$ is also upper (lower) triangular Toeplitz. *Proof.* The general form for the inverse of a Toeplitz martix given in [45] is factored into Toeplitz matrices as

$$pT^{-1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_N & \cdots & a_1 & 1 \end{bmatrix} \begin{bmatrix} 1 & b_1 & \cdots & b_N \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b_1 \\ 0 & \cdots & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & \cdots & 0 \\ b_N & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ b_1 & \cdots & b_N & 0 \end{bmatrix} \begin{bmatrix} 0 & a_N & \cdots & a_1 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_N \\ 0 & \cdots & 0 & 0 \end{bmatrix} \tag{217}$$

where

$$T\mathbf{a} = \begin{bmatrix} t_0 & t_{-1} & \cdots & t_{-N} \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{-1} \\ t_N & \cdots & t_1 & t_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} \rho \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{218}$$

and

$$T\mathbf{b} = \begin{bmatrix} t_0 & t_{-1} & \cdots & t_{-N} \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{-1} \\ t_N & \cdots & t_1 & t_0 \end{bmatrix} \begin{bmatrix} b_N \\ \vdots \\ b_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \rho \end{bmatrix}. \tag{219}$$

If $T$ is upper triangular, $t_i = 0 \quad i = -1 \ldots - N$, and since $T^{-1}$ is also upper triangular $\mathbf{a}$ is simply the first column of the inverse, normalized as $\mathbf{a} = \begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix}^T$. Substituting this in Equation 217 gives

$$pT^{-1} = \begin{bmatrix} 1 & b_1 & \cdots & b_N \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b_1 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \tag{220}$$

Thus $T^{-1}$ is Toeplitz. The lower triangular case follows similarly.

**Theorem.** The operations in Equation 105 commute. *Proof.* Let $Z$ be the $N \times N - p$ matrix consisting of the $N - p$ identity matrix augmented with $p$ rows of zeros and $\bar{Z}$ be the $N \times N$ matrix consisting of $Z$ augmented with $p$ columns of zeros. Then $\bar{A}$ is the $N \times N$ lower triangular Toeplitz matrix such $A = \bar{A}Z$. The inverse of the covariance matrix $R^{-1} = (BP^{-1}B^*)^{-1}$ is approximated by zeroing the last $p$ rows of $B$ of with $\bar{K} = \bar{Z}B$ and resizing $B$ with $\tilde{K} = Z^*B$. The neglected terms constitute an approximation with $O(p/N)$ error. The operations in Equation 105 are approximated as

$$(A^*(BP^{-1}B^*)^{-1}A)^{-1} \simeq (A^*(\bar{K}\bar{K}^*)^{+}A)^{-1} \tag{221}$$

For $q < p$ , excepting zeroed elements $\bar{K} = \bar{Z}B$ is lower triangular Toeplitz. Then excepting zeroed elements, matrices $(\bar{K})^+$ and $\tilde{A} = \bar{Z}\bar{A}$ are lower triangular Toeplitz and the same rank and thus

they commute.

$$
\begin{aligned}
(A^*(\bar{K}\bar{K}^*)^+A)^{-1} &= (Z^*\bar{A}^*\bar{Z}(\bar{K}\bar{K}^*)^+\bar{Z}\bar{A}Z)^{-1} && (222)\\
&= (Z^*\tilde{A}^*(\bar{K}^*)^+(\bar{K})^+\tilde{A}Z)^{-1} && (223)\\
&= (Z^*(\bar{K}^*)^+\tilde{A}^*\tilde{A}(\bar{K})^+Z)^{-1} && (224)\\
&= \tilde{K}(\tilde{A}^*\tilde{A})^+\tilde{K}^* && (225)
\end{aligned}
$$

since $\bar{Z}Z = Z$ and $Z^+ = Z^*$. The $(\tilde{A}^*\tilde{A})^+$ term is also approximated with $O(p/N)$ error by $(\bar{A}^*\bar{A})^{-1}$ or $Z(A^*A)^{-1}Z^*$.

## *Appendix B. Consistent 1-D Noise Estimates*

Consistent estimates of stationary, autoregressive (AR), and autoregressive-moving average (ARMA) are attained through application of the following theorems.

**Theorem** (stationary). A consistent estimate of the noise covariance matrix $R_{\mathbf{ww}}$ of stationary noise sequence is attained from the difference between the observed data and the maximum likelihood estimate of the signal. *Proof.* Given the Toeplitz structure of the stationary noise covariance matrix $R_{\mathbf{ww}}$ and a consistent estimator of the elements of the covariance matrix the autocorrelation coefficients, all that remains to be shown is that the residuals of the maximum likelihood estimate are a consistent estimate of the noise sequence. The estimate is

$$\hat{\mathbf{w}} = \mathbf{y} - \hat{G}\hat{\mathbf{s}} \tag{226}$$

where $\hat{G}\hat{\mathbf{s}}$ is the ML estimate of the signal. By the triangle inequality

$$\begin{aligned} \|\hat{\mathbf{w}} - \mathbf{w}\| &= \|\mathbf{y} - \hat{G}\hat{\mathbf{s}} - (\mathbf{y} - G\mathbf{s})\| \\ &\leq \|\mathbf{y} - \mathbf{y}\| + \|G\mathbf{s} - \hat{G}\hat{\mathbf{s}}\| \\ &= \|G\mathbf{s} - \hat{G}\hat{\mathbf{s}}\| \end{aligned} \tag{227}$$

Since $\|\cdot\| > 0$, squaring both sides and taking the limit of the expected value gives

$$\lim_{N\to\infty} \mathrm{E}\{\|\hat{\mathbf{w}} - \mathbf{w}\|_2^2\} \leq \lim_{N\to\infty} \mathrm{E}\{\|G\mathbf{s} - \hat{G}\hat{\mathbf{s}}\|_2^2\}. \tag{228}$$

And since the ML estimates of the signal are consistent

$$\lim_{N\to\infty} \mathrm{E}\{\|\hat{\mathbf{w}} - \mathbf{w}\|_2^2\} = 0. \tag{229}$$

**Theorem** (AR). Consistent estimates of a AR($p$) noise process are given by the approximating the maximum likelihood minimization as shown in Appendix A. Asymptotically, the same minimization gives consistent estimates of the AR noise process. *Proof.* For a fixed signal the ML

estimate of the noise is given by

$$\min_{R_{\mathbf{ww}}} \frac{1}{N} \ln |\mathbf{y}^* A(A^* R_{\mathbf{ww}} A)^{-1} A^* \mathbf{y} R_{\mathbf{ww}}| = \min_{R_{\mathbf{ww}}} \frac{1}{N} \ln |\mathbf{y}^* A(A^* R_{\mathbf{ww}} A)^{-1} A^* \mathbf{y} I| + \frac{1}{N} \ln |R_{\mathbf{ww}}|. \quad (230)$$

Taking the limit and introducing the reflection coefficient decomposition of $\ln |R_{\mathbf{ww}}|$ from [51] where $|k_n| < 1$ gives

$$\begin{aligned} & \min_{R_{\mathbf{ww}}} \ln \mathbf{y}^* A(A^* R_{\mathbf{ww}} A)^{-1} A^* \mathbf{y} - \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{p} n(1 - k_n) \\ = & \min_{R_{\mathbf{ww}}} \mathbf{y}^* A(A^* R_{\mathbf{ww}} A)^{-1} A^* \mathbf{y}. \end{aligned} \quad (231)$$

**Theorem** (ARMA). Consistent estimates of a ARMA($p$) noise process are also given by the approximating the maximum likelihood minimization as shown in Appendix A. The AR and MA parts of the noise can be estimated separately when the AR part of the noise process contains most of the noise energy (i.e. the noise poles are close to the unit circle in the complex plane, and the noise poles are separated in frequency from the moving average zeros). *Proof.* The AR process is estimated first. Then the AR theorem may be applied. The stationary theorem may then be applied to the whitened residuals of the noise process. $\hat{\mathbf{w}}_{MA} = L_{AR}^{+}(\mathbf{y} - \hat{G}\hat{\mathbf{s}})$ where $R_{AR} = L_{AR}^{*} L_{AR}$ and $(\cdot)^{+}$ indicates pseudoinverse.

## *Appendix C. A Bound on the MSE of Estimates*

A useful measure of how well the estimates, $\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 & \hat{\theta}_2 & \hat{\theta}_p \end{bmatrix}^T$, match the parameters, $\theta$, of a model is Mean Square Error (MSE),

$$\text{MSE} = (\hat{\theta}_1 - \theta_1)^2. \tag{232}$$

When the error covariance matrix is defined as

$$C = \text{E}\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\}, \tag{233}$$

the MSE of the $i^{th}$ parameter is the $i^{th}$ diagonal element of the matrix $C$ , $(C)_{ii}$. And when the estimate is unbiased, $\text{E}\{\hat{\theta}\} = \theta$, the error covariance matrix is the covariance matrix of the estimator,

$$\begin{aligned} C &= \text{E}\{(\hat{\theta} - \mathbf{E}\{\hat{\theta}\} + \mathbf{E}\{\hat{\theta}\} - \theta)(\hat{\theta} - \mathbf{E}\{\hat{\theta}\} + \mathbf{E}\{\hat{\theta}\} - \theta)^T\} & (234)\\ &= \text{E}\{(\hat{\theta} - \mathbf{E}\{\hat{\theta}\})(\hat{\theta} - \mathbf{E}\{\hat{\theta}\})^T\} + (\mathbf{E}\{\hat{\theta}\} - \theta)(\mathbf{E}\{\hat{\theta}\} - \theta)^T & (235)\\ &= \text{E}\{(\hat{\theta} - \mathbf{E}\{\hat{\theta}\})(\hat{\theta} - \mathbf{E}\{\hat{\theta}\})^T\}. & (236) \end{aligned}$$

In Maximum Likelihood (ML) estimation, the estimates of the parameters are given by the zeros of the score function $\mathbf{s}(\theta, \mathbf{x})$, the derivative of the probability density, $f$, of the observed data, $\mathbf{x}$, with respect to the parameters $\theta$,

$$\mathbf{s}(\theta, \mathbf{x}) = \frac{\partial}{\partial \theta} \ln f_\theta(\mathbf{x}) = 0. \tag{237}$$

When the ML estimate exists, it is unbiased and a bound on the ML MSE is the Cramér-Rao Bound (CRB). The following two facts reveal this bound.

1) The score function is identically correlated with the error [66]. Since the estimate is unbiased

$$\text{E}\{\hat{\theta} - \theta\} = \mathbf{0} \tag{238}$$

or equivalently

$$\int d\mathbf{x} f_\theta(\mathbf{x})(\hat{\theta} - \theta)^T = \mathbf{0}^T \tag{239}$$

Under regularity conditions, the derivative with respect to $\theta$ of this equation is given by

$$\int d\mathbf{x} \frac{\partial}{\partial \theta} f_\theta(\mathbf{x})(\hat{\theta} - \theta)^T - \int d\mathbf{x} f_\theta(\mathbf{x}) I = \mathbf{0}^T, \tag{240}$$

or since $\frac{\partial}{\partial \theta} f_\theta(\mathbf{x}) = f_\theta(\mathbf{x}) \frac{\partial}{\partial \theta} \ln f_\theta(\mathbf{x})$ and $\int d\mathbf{x} f_\theta(\mathbf{x}) = 1$,

$$\int d\mathbf{x} f_\theta(\mathbf{x}) \frac{\partial}{\partial \theta} \ln f_\theta(\mathbf{x})(\hat{\theta} - \theta)^T = I. \tag{241}$$

Which is the first fact,

$$\mathrm{E}\{\mathbf{s}(\theta, \mathbf{x})(\hat{\theta} - \theta)^T\} = I. \tag{242}$$

2) A matrix inequality [51]. The covariance matrix of the random vector $\mathbf{y} - R_{yz} R_{zz}^{-1} \mathbf{z}$ is positive semidefinite,

$$\begin{aligned} \mathrm{E}\{(\mathbf{y} - R_{yz} R_{zz}^{-1} \mathbf{z})(\mathbf{y} - R_{yz} R_{zz}^{-1} \mathbf{z})^T\} &= R_{yy} - 2R_{yz} R_{zz}^{-1} R_{yz}^T + R_{yz} R_{zz}^{-1} R_{zz} R_{zz}^{-1} R_{yz}^T \\ &= R_{yy} - R_{yz} R_{zz}^{-1} R_{yz}^T \geq 0. \end{aligned} \tag{243}$$

Let $\mathbf{y} = \hat{\theta} - \theta$ and $\mathbf{z} = \mathbf{s}(\theta, \mathbf{x})$ then $R_{yz} = I$ from the first fact and from the second fact the CRB is

$$C \geq (\mathrm{E}\{\mathbf{s}(\theta, \mathbf{x}) \mathbf{s}(\theta, \mathbf{x})^T\})^{-1} = J^{-1}(\theta), \tag{244}$$

where $J(\theta) = \mathrm{E}\{\mathbf{s}(\theta, \mathbf{x}) \mathbf{s}(\theta, \mathbf{x})^T\}$ is the Fisher Information matrix. Then the MSE of the $i^{th}$ parameter is bounded by the $i^{th}$ diagonal term of the inverse of the FIM.

$$\mathrm{E}(\hat{\theta}_i - \theta_i)^2 \geq (J^{-1})_{ii}. \tag{245}$$

## *Appendix D. Cramér-Rao Bound for Deterministic Exponentials*

This appendix describes the Cramér-Rao (CR) bound for damped exponentials in colored noise. The bounds are computed by inverting the Fisher information matrix (FIM). The general form for the $kl^{th}$ element of the FIM for a complex circular Gaussian process [71] is described in terms of the mean $G\mathbf{s}$ and covariance $\sigma^2 R_{\mathbf{ww}}$ of the damped exponential in colored noise process as

$$J_{kl} = 2\mathrm{Re}(\frac{\delta G\mathbf{s}}{\delta\theta_k}\sigma^{-2}R_{\mathbf{ww}}^{-1}\frac{\delta G\mathbf{s}}{\delta\theta_l}) + \mathrm{tr}(\sigma^{-2}R_{\mathbf{ww}}^{-1}\frac{\delta\sigma^2 R_{\mathbf{ww}}}{\delta\theta_k}\sigma^{-2}R_{\mathbf{ww}}^{-1}\frac{\delta\sigma^2 R_{\mathbf{ww}}}{\delta\theta_l}), \tag{246}$$

where the parameters are represented in the real vector

$$\theta = [\ \sigma^2 \quad |\eta|^T \quad \arg(\eta)^T \quad |\mathbf{s}|^T \quad \arg(\mathbf{s})^T \quad |\lambda|^T \quad \arg(\lambda)^T]^T\ , \tag{247}$$

whose $k^{th}$ element is $\theta_k$. As is the case for undamped exponentials [16] the estimation bounds for the exponentials and the colored noise are decoupled and the FIM has the structure,

$$J = \begin{bmatrix} J(\sigma^2,\eta) & \mathbf{0} \\ \mathbf{0} & J(\mathbf{s},\lambda) \end{bmatrix}, \tag{248}$$

where

$$J(\sigma^2,\eta)_{kl} = \frac{1}{\sigma^4}\mathrm{tr}(R_{\mathbf{ww}}^{-1}\frac{\delta\sigma^2 R_{\mathbf{ww}}}{\delta\theta_k}R_{\mathbf{ww}}^{-1}\frac{\delta\sigma^2 R_{\mathbf{ww}}}{\delta\theta_l}), \tag{249}$$

$$J(\mathbf{s},\lambda)_{kl} = \frac{2}{\sigma^2}\mathrm{Re}(\frac{\delta G\mathbf{s}}{\delta\theta_k}R_{\mathbf{ww}}^{-1}\frac{\delta G\mathbf{s}}{\delta\theta_l}). \tag{250}$$

This decoupling is due to the Wold decomposition [51] which shows that any regular wide sense stationary random process can be separated into a regular random process and a purely deterministic process. Specific bounds for an AR noise process and the $J(\sigma^2,\eta)$ block of the FIM are developed in [16]. The structure of the damped exponential block has also been developed [67], [10] for the case of white noise and colored noise, respectively.

Compactly written the damped exponential bounds are

$$J(\mathbf{s},\lambda) = \frac{2}{\sigma^2}\mathrm{Re}(F^* R_{\mathbf{ww}}^{-1} F) \tag{251}$$

where the $N \times 4p$ matrix $F$ is

$$F \equiv [\ G\Upsilon \quad jG\Sigma \quad DG\Sigma\Lambda^{-1} \quad jDG\Sigma\ ] \tag{252}$$

where

$$\Upsilon \equiv \text{diag}\{\ \exp(j\arg(s_1)) \quad \exp(j\arg(s_2)) \quad \cdots \quad \exp(j\arg(s_p))\ \}, \tag{253}$$

$$\Sigma \equiv \text{diag}\{\ s_1 \quad s_2 \quad \cdots \quad s_p\ \}, \tag{254}$$

$$\Lambda \equiv \text{diag}\{\ |\lambda_1| \quad |\lambda_2| \quad \cdots \quad |\lambda_p|\ \}, \tag{255}$$

$$D \equiv \text{diag}\{\ 0 \quad 1 \quad \cdots \quad N\ \}. \tag{256}$$

The CR bounds are then given by $\mathrm{E}(\hat{\theta}_k - \theta_k)^2 \geq (J^{-1})_{kk}$.

## *Appendix E. Method of Direction Estimation (MODE)*

The MODE algorithm is described in [71] [69], and [82]. MODE minimizes

$$\min_{A} \operatorname{tr}\left(A(A^*A)^{-1}A^*VWV^*\right) \tag{257}$$

where the eigendecomposition of the data matrix is $YY^* = V(D - \sigma^2 I_{p\times p})V^* + \sigma^2 I_{M\times M}$ and the columns $V$ are the eigenvectors of $Y$ associated with the $p$ largest eigenvalues (the elements of the diagonal matrix $D$), $\sigma^2$ is the average of the remaining eignevalues and the optimal weighting [50] is $W = (D - \sigma^2 I_{p\times p})^2 D^{-1}$. The minimization in equation 257 is computed in a similar manner to IQML as

$$\min_{\mathbf{a}_i} \mathbf{a}_i^* F^* F \mathbf{a}_i. \tag{258}$$

where $F = [\ CF_1 \qquad CF_p\ ]^T$ and $C = (A^*A)_{i-1}^{-1/2}$

$$F_n{}^T = \begin{bmatrix} V_{p+1,n} & \cdots & V_{1,n} \\ V_{P+2,n} & \cdots & V_{2,n} \\ \vdots & & \vdots \\ V_{M,n} & \cdots & V_{M-p,n} \end{bmatrix}$$

and $V_{kl}$ is the $kl^{th}$ element of $VW^{1/2}$. Essentially a data matrix similar to $Y$ is formed from each eigenvector in $V$ weighted by $W^{1/2}$ and these data matrices are stacked (for simultaneous minimization).

## *Appendix F. A lower memory method for implementing 2-D IQML*

In the original version of 2D-IQML the following procedure produce the kernel that is used to predict the 2-D frequencies:

$$T = (WW^*)^+\tilde{Y}$$

$$K = \tilde{Y}^*T.$$

Since $W$ is $((2M_1-1)(M_2-p)+M_1-1)\times M_1M_2$ this requires the SVD and pseudoinverse of a $((2M_1-1)(M_2-p)+M_1-1)\times((2M_1-1)(M_2-p)+M_1-1)$ matrix. The same kernel can be produced with less memory by taking advantage of the smaller column width of $W$ and the fact that $(WW^*)^+ = ((W^*W)^+W^*)^*(W^*W)^+W^*$ which give us the following procedure:

$$S = (W^*W)^+W^*\tilde{Y}$$

$$K = S^*S$$

We can see that the results are equivalent by expanding both equations. In the first case

$$K = \tilde{Y}^*((UDV^*)(UDV^*)^*)^+\tilde{Y} = \tilde{Y}^*(UDV^*VD^*U^*)^+\tilde{Y}$$

$$= \tilde{Y}^*(UDD^*U^*)^+\tilde{Y} = \tilde{Y}^*U(DD^*)^+U^*\tilde{Y}$$

In the second case

$$S = ((UDV^*)^*(UDV^*))^+(UDV^*)^*\tilde{Y} = (VD^*U^*UDV^*)^+VD^*U^*\tilde{Y}$$

$$= (VD^*DV^*)^+VD^*U^*\tilde{Y} = V(D^*D)^+V^*VD^*U^*\tilde{Y} = V(D^*D)^+D^*U^*\tilde{Y}.$$

And

$$K = S^*S = (V(D^*D)^+D^*U^*\tilde{Y})^*V(D^*D)^+D^*U^*\tilde{Y}$$

$$= \tilde{Y}^*UD(D^*D)^{+*}V^*V(D^*D)^+D^*U^*\tilde{Y} = \tilde{Y}^*UD(D^*D)^{+*}(D^*D)^+D^*U^*\tilde{Y}$$

where all the inner matrices are diagonal and commute. Thus

$$K = \tilde{Y}^* U (D^* D)^{+*} (D^* D)^{+} (D^* D) U^* \tilde{Y} = \tilde{Y}^* U (D D^*)^{+} U^* \tilde{Y}$$

or the same result as the first case.

## *Bibliography*


1. D. J. Andersh, M. Hazlett, S. W. Lee, D. D. Reeves, D. P. Sullivan and Y. Chu, "XPATCH: A high frequency electromagnetic scattering prediction code and environment for complex three dimensional objects," *IEEE Ant., Prop. Mag.*, vol. 36, no. 2, pp. 65-69, Feb. 1994.
2. D. V. Bhaskar Rao and S. Y. Kung, "A state-space approach for the 2-D harmonic retrieval problem," *Proc. IEEE ICASSP 1994*, pp. 4.10.1-4.10.4.
3. J. F. Böhme, "Estimation of source parameters by maximum likelihood and nonlinear regression," in *Proc. IEEE ICASSP 1984*, pp. 7.3.1–7.3.4, 1984.
4. J. F. Böhme, "Estimation of spectral parameters of correlated signals in wavefields," *Signal Processing*, 10:329-337, 1986.
5. J. F. Böhme and D. Kraus, "On least squares methods for direction of arrival estimation in the presence of unknown colored noise," in *Proc. IEEE ICASSP 1988*, pp. 2833-2835.
6. Y. Bresler and A. Macovski, "Exact maximum likelihood estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-34, pp. 1081-1089, Oct. 1986.
7. G. Box and G. M. Jenkins, *Time Series Analysis Forecasting and Control*, San Francisco: Holdenday, 1970.
8. C. Cafforio, C. Prati, and E. Rocca, "SAR data focusing using seismic migration techniques," *IEEE Trans. Ant., Elec. Sys.*, vol. 27, no. 2, pp. 194-206, Mar. 1991.
9. M. P. Clark, "Estimation techniques for sensor array processing," Ph.D. Thesis, University of Colorado, 1992.
10. M. P. Clark and L. L. Scharf, "Two-dimensional modal analysis based on maximum likelihood," *IEEE Trans. Signal Processing*, vol. 42, no. 6, pp. 1443-1451, June 1994.
11. M. P. Clark, "Cramér-Rao bounds for two-dimensional deterministic modal analysis," in *Proc. 27th Asilomar Conf. on Sig., Sys., Comp.*, Oct. 1993.
12. M. P. Clark and L. L. Scharf, "A maximum likelihood estimation technique for spatial-temporal modal analysis," in *Proc. 25th Asilomar Conf. on Sig., Sys., and Comp.*, Oct. 1991.
13. L. Cohen, "Preface to special issue on time-frequency analysis," *Proc. IEEE*, vol. 84, no. 9, pp. 1197-1198, Sep. 1996.
14. D. E. Dudgeon and Mersereau, *Multidimensional Digital Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1984.
15. P. H. Eichel, D. C. Ghiglia, and C. V. Jakowatz, "A speckle processing method for synthetic aperture radar phase correction", Sandia National Laboratories Report 1988.
16. J. M. Francos and B. Friedlander, "Bounds for estimation of complex exponentials in unknown colored noise," *IEEE Trans. Signal Processing*, vol. 43, no. 9, pp. 2176-2185, Sep. 1992.
17. B. Friedlander and A. J. Weiss, "On direction finding with unknown noise covariance," in *Proc. 28th Asilomar Conf. on Sig., Sys., Comp.*, Oct. 1994.
18. I. J. Gupta, "High-resolution radar imaging using 2-D linear prediction," *IEEE Trans. Ant. Prop.*, vol. 42, no. 1, pp. 31-37, Jan. 1995.
19. G. H. Golub and C. F. Van Loan, *Matrix Computations*, Baltimore: John Hopkins University Press, 1989.
20. H. Harada, D. J. Wall, T. Takenaka, and M. Tanaka, "Conjugate gradient method applied to inverse scattering problem," *IEEE Trans. Ant. Prop.*, vol. 43, no. 8, pp. 784-792, Aug. 1995.

21. E. J. Hannan, "The estimation of frequency," *J. Appl. Prob.*, pp. 510-519, 1973.

22. R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge: Cambridge University Press, 1985.

23. Y. Hua, "The most efficient implementation of the IQML algorithm," *IEEE Trans. Signal Processing*, vol. 42, no. 8, pp. 2203-2204, Aug. 1994.

24. Y. Hua, E. Baqui, Y. Zhu, and D. Heilbronn, "Imaging of point scatters from step-frequency ISAR data," *IEEE Trans. Aero., Elec. Sys.*, vol. 29, no. 1, pp. 195-204, Jan. 1993.

25. Y. Hua, "Estimation two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Trans. Signal Processing*, vol 40, no. 9, pp. 2267-2280, Sep. 1992.

26. A. G. Jaffer, "Maximum likelihood direction finding of stochastic sources: a separable solution," in *Proc. ICASSP 88,* vol 5, pp. 2893-2896, New York, Ney York, Apr. 1988.

27. D. H. Johnson and D. E. Dudgeon, *Array Signal Processing Concepts and Techniques*, Englewood Cliffs, NJ: Prentice Hall, 1993.

28. A. K. Jain, *Fundamentals of Digital Image Processing*, Englewood Cliffs, NJ: Prentice Hall, 1989.

29. S. M. Kay and V. Nagesha, "Maximum likelihood estimation of signals in autoregressive noise," *IEEE Trans. Signal Processing*, vol. 41, no. 1, pp. 88-101, Jan. 1993.

30. S. M. Kay, *Modern Spectral Estimation Theory and Applications*, Englewood Cliffs, NJ: Prentice Hall, 1988.

31. S. M. Kay and V. Nagesha, "Estimation for processes with mixed spectra," in *Proc. IEEE ICASSP*, 1993, pp. IV-232.

32. H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Processing Magazine* vol. 13, no. 4, pp. 67-94, Jul. 1996.

33. R. Kumaresan, L. L. Sharf and A. K. Shaw, "An algorithm for pole-zero modeling and spectral analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-34, pp. 637-640, June 1986.

34. R. Kumaresan and A. K. Shaw, "Superresolution by structured matrix approximation," *IEEE Trans. Aero., Elec. Sys.*, vol. 36, no. 1, pp. 34-44, Jan. 1988.

35. R. Kumarsan, "Spectral Analysis," *Digital Signal Processing*, S. K. Mitra and J. F. Kaiser, eds. New York: Wiley, 1993.

36. J. P. Le Cadre, "Parametric methods for spatial signal processing in the presence of unknown colored noise fields," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 7, pp. 965-983, Jul. 1989.

37. P. Lancaster, and K. Salkauskas, *Curve and Surface Fitting*, London: Academic Press, 1986.

38. J. Li, P. Stoica, and D. Zheng, "An efficient algorithm for two-dimensional frequency estimation," *Multidimensional Systems and Signal Processing*, vol. 7, no. 2, pp. 151-178, Apr. 1996.

39. J. Li, P. Stoica, "Efficient mixed spectrum estimation with applications to target feature extraction," *IEEE Trans. Signal Processing*, vol. 44, no. 2, pp. 281-295, Feb. 1996.

40. J. Li, P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Trans. Signal Processing*, vol. 44, no. 6, pp. 1469-1484, Jun. 1996.

41. J. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Economics*, New York: Wiley, 1988.

42. D. L. Mensa, *High Resolution Radar Cross-Section Imaging*, Boston: Artech House, 1981.

43. D. L. Moffatt and R. K. Mains, "Detection and discrimination of radar target," *IEEE Trans. Ant., Prop.*, vol. 23, no. 5, pp. 358-367, May 1975.

44. D. L. Moffatt and C. M. Rhoads, "Radar identification of Naval Vessels," *IEEE Trans. Aero., Elec. Sys.*, vol. 18, no. 2, pp. 182-187, Mar. 1982.

45. S. L. Marple, *Digital Spectral Analysis with Applications*, Englewood Cliffs, NJ: Prentice Hall, 1987.

46. V. Nagesha and S. M. Kay, "On frequency estimation with the IQML algorithm," *IEEE Trans. Signal Processing*, vol. 42, no. 9, pp. 2507-2513, Sep. 1994.

47. V. Nagesha and S. M. Kay, "Maximum likelihood estimation for array processing in colored noise," *IEEE Trans. Signal Processing*, vol. 44, no. 2, pp. 169-180, Feb. 1996.

48. V. Nagesha and S. M. Kay, "Spectral analysis based on the canonical autoregressive decomposition," *IEEE Trans. Signal Processing*, vol. 44, no. 7, pp. 1719-1733, Jul. 1996.

49. R. H. Norden, "A survey of maximum likelihood estimation," *Interantional Statistical Review*, vol. 40, pp. 329-354,1972.

50. B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample maximum likelihood techniques for parametric estimation and detection in array processing," *Radar Array Processing*, S. Haykin, ed. New York: Springer-Verlag, 1993.

51. B. Porat, *Digital Processing of Random Signals*, Englewood Cliffs, NJ: Prentice Hall, 1994.

52. M. P. Pepin and M. P. Clark, "On the performance of several 2-D harmonic retrieval techniques," in *Proc. 28th Asilomar Conf. on Sig., Sys., Comp.*, Oct. 1994.

53. M. P. Pepin and M. P. Clark, "On the applicability of 2-D damped exponential models to synthetic aperture radar," in *Proc. IEEE ICASSP 1995*, May 1995.

54. V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. R. Astron. Soc.*, vol.33, pp. 347-366, 1973.

55. R. Prony, *Essai expérimental et analytique*, vol. 1, no. 2, pp. 24-76, L'Ecole Polytechnique, Paris, 1795.

56. R. A. Roberts, and C. T. Mullis, *Digital Signal Processing*, Reading Mass.: Addison-Wesley, 1987.

57. R. H. Roy, A. Paulraj, and T. Kailath, "ESPRIT - a subspace rotational approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-34, pp. 1340-1342, Oct. 1986.

58. J. J. Sacchini, "Development of two-dimensional parametric radar signal modeling and estimation techniques with application to target identification," Ph.D. Thesis, Ohio State University, 1992.

59. J. J. Sacchini, W. M. Steedly, and R. L. Moses, "Two-dimensional Prony modeling and parameter estimation," *IEEE Trans. Signal Processing*, vol. 41, no. 11, pp. 3127-3137 Nov. 1993.

60. D. T. Sandwell, "Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data," Geophysical Research Letters, no. 2, pp. 139-142, 1987.

61. R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Ant., Prop.*,AP-34, no. 3, pp. 276-280, Mar. 1986.

62. A. Shaw, "Maximum likelihood estimation of multiple frequencies with constraints to guarantee unit circle roots," *IEEE Trans. Signal Processing*, vol. 43, no. 3, pp. 796-799, Mar. 1995.

63. M. Soumekh, M., *Fourier Array Imaging*, Englewood Cliffs, NJ: Prentice Hall, 1994.

64. M. Soumekh, M., "Reconnaissance with ultra wideband UHF synthetic aperture radar", *IEEE Sig. Proc. Mag.*, vol. 12, no. 4, pp. 21-40, Jul. 1995.

65. M. Soumekh, "A system model and inversion for synthetic aperture radar imaging", *IEEE Trans. Image Proc.*, vol. 1, no. 1, pp. 64-76, Jan. 1992.

66. L. L. Scharf, *Statistical Signal Processing*, Reading, Mass: Addison-Wesley, 1991.

67. W. D. Steedy and R. L. Moses, "The Cramér-Rao bound for pole and amplitude coefficients of damped exponential signals in noise," *IEEE Trans. Signal Processing*, vol. 41, no. 3, pp. 1305-1318, Mar. 1993.

68. P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and the Cramér-Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 5, pp. 720-741, May. 1989.

69. P. Stoica and K. Sharman, "Novel eigenanalysis method for direction estimation," IEE Proceeding, February 1990.

70. P. Stoica and A. Nehorai, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 7, pp. 1132-1143, Jul. 1990.

71. P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 10, pp. 1783-1795, Oct. 1990.

72. P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and the Cramér-Rao bound: further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 12, pp. 2140-2150, Dec. 1990.

73. P. Stoica, M. Viberg, and B. Ottersten, "Instrumental variable approach to array processing in spatially correlated noise fields," *IEEE Trans. Signal Processing*, vol. 42, no. 1, pp. 121-133, Jan 1994.

74. C. W. Therrien, *Digital Random Signals and Statistical Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1992.

75. P. J. Tourtier and L. L. Scharf, "Maximum Likelihood Identification of Correlation Matrices for Estimation of Power Spectra at Arbitrary Resolutions," in *Proc. IEEE ICASSP 1987*, pp. 2066-2069.

76. D. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, no. 9, pp. 975-990, Sep. 1980.

77. M. J. Turmon and M. I. Miller, "Maximum likelihood estimation of complex sinusoids and Toeplitz covariances," *IEEE Trans. Signal Processing*, vol. 42, no. 5, pp. 1074-1086, May 1994.

78. M. Viberg, "Sensitivity of parametric direction finding to colored noise fields and undermodeling," *IEEE Trans. Signal Processing*, vol. 40, no. 1, pp. 245-249, Jan. 1992.

79. M. Viberg, B. Ottersten, and A. Nehorai, "Performance analysis of direction finding with large arrays and finite data," *IEEE Trans. Signal Processing*, vol. 43, no. 2, pp. 469-477, Feb. 1996.

80. M. Wax, *Detection and estimation of superimposed signals,* PhD thesis, Stanford Univ., Stanford, CA, Mar. 1985.

81. M. Wax, "Detection and localization of multiple sources in noise with unknown covariance," *IEEE Trans. Signal Processing*, vol. 40, no. 1, pp. 245-249, Jan. 1992.

82. A. J. Weiss, B. Friedlander, and P. Stoica, "Direction of arrival estimates using MODE with interpolated arrays," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 296-300, Jan. 1995.

83. N. Wiener, "Generalized harmonic analysis," *Acta Math.*, vol. 55, pp. 117-258, 1930.

84. H. Wold, *A Study in the Analysis of Stationary Time Series*, Uppsala, Sweden: Almqvist & Wiksell, 1938.

85. Q. Wu, K. M. Wong, and J. P. Reilly, "Maximum likelihood direction finding in unknown noise environments," *IEEE Trans. Signal Processing*, vol. 42, no. 4, pp. 980-983, Apr. 1994

86. H. Ye and R. DeGroat, "Maximum likelihood DOA estimation and asymptotic Cramér-Rao bounds for additive unknown colored noise," *IEEE Trans. Signal Processing*, vol. 43, no. 4, pp. 938-949, Apr. 1995.

87. G. U. Yule, "On a method of investigating periodicities in disturbed with special reference to Wolfer's sun-spot numbers," *Philos. Trans. R. Soc. London, Ser. A*, vol. 226, pp. 276-298, 1927.

88. I. Ziskind. and M. Wax, "Maximum likelihood DOA localization of multiple sources by alternating projection," *IEEE Trans. Signal Processing*, vol. 36, no. 10, pp. 1553-1560, Oct. 1988.

89. G. Xu and T. Kailath, "Fast Subspace Decomposition," *IEEE Trans. Signal Processing*, vol. 42, no. 3, pp. 539-551, Mar. 1994.

## *Vita*

Matthew P. Pepin [redacted] He graduated from Faifax High School, Fairfax, Virginia in 1977. In 1981, he received his B.S.E.E. degree from the University of Virginia and received an ROTC commission in the United States Air Force. After receiving his M.E.E.E from the University of Virginia in 1982 he was assigned to Eglin AFB, Florida with the infrared avionics systems test branch. The tests he conducted included airborne infrared flares and aircraft measurements, the Airfield Damage Assessment System and the Infrared Search and Track System, as well as European deployment to measure NATO aircraft infrared signatures. Captain Pepin and his wife Fay were married in Niceville Florida in 1986. In 1986 following German Language training at the Defense Language Institute in Monterey, California, he was selected as an Engineering Exchange Officer with Dornier GmbH in Friedrichshafen, West Germany where he helped develop a laser scanner for a robotic system. After returning from Germany in 1988 he was assigned to special programs and avionics systems development for the B-2 System Program Office. At the B-2 SPO he led the engineering development effort for special systems, integration of primary avionics functions and computer programs for the B-2. He chaired the B-2 development team for the Tactical Situation Display. In 1993, he was selected to attend the Air Force Institute of Technology (AFIT) as a PhD candidate. He has two daughters both born during his studies at AFIT in [redacted] and in [redacted]

[redacted]

# REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188



| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | September 1996 | PhD Dissertation |

**4. TITLE AND SUBTITLE**
Maximum Likelihood Estimation of Exponentials in Unknown Colored Noise for Target Identification in Synthetic Aperture Radar Images

**5. FUNDING NUMBERS**

**6. AUTHOR(S)**
Matthew P. Pepin, Maj, USAF

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
AFIT/ENG
Air Force Institute of Technology
2950 P. Street
Wright-Patterson Air Force Base, OH 45433-6583

**8. PERFORMING ORGANIZATION REPORT NUMBER**
AFIT/DS/ENG/96-09

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
WL/AARA
Wright Laboratories
Wright-Patterson Air Force Base, OH 45433

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**
Approved for Public Release; Distribution Unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*

This dissertation develops techniques for estimating exponential signals in unknown colored noise. The Maximum Likelihood (ML) estimators of the exponential parameters are developed. Techniques are developed for one and two-dimensional exponentials, for both the deterministic and stochastic ML model. The techniques are applied to Synthetic Aperture Radar (SAR) data whose point scatterers are modeled as damped exponentials. These estimated scatterer locations (exponentials frequencies) are potential features for model-based target recognition. The estimators developed in this dissertation may be applied with any parametrically modeled noise having a zero mean and a consistent estimator of the noise covariance matrix. ML techniques are developed for a single instance of data in colored noise which is modeled in one dimension as 1) stationary noise, 2) autoregressive (AR) noise and 3) autoregressive moving-average (ARMA) noise and in two dimensions as 1) stationary noise, and 2) white noise driving an exponential filter. The classical ML approach is used to solve for parameters which can be decoupled from the estimation problem. The remaining nonlinear optimization to find the exponential frequencies is then solved by extending white noise ML techniques to colored noise. In the case of deterministic ML, the computationally efficient, one and two-dimensional Iterative Quadratic Maximum Likelihood (IQML) methods are extended to colored noise. In the case of stochastic ML, the one and two-dimensional Method of Direction Estimation (MODE) techniques are extended to colored noise. Simulations show that the techniques perform close to the Cramér-Rao bound when the model matches the observed noise.

| 14. SUBJECT TERMS | 15. NUMBER OF PAGES |
|---|---|
| Exponential Estimation, Synthetic Aperture Radar | 170 |
| | 16. PRICE CODE |

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| UNCLASSIFIED | UNCLASSIFIED | UNCLASSIFIED | UL |